DOCUMENT RESUME

ED 304 446 TM 012 667

AUTHOR Ohlsson, Stellan
TITLE Trace Analysis and Spatial Reasoning: An Example of Intensive Cognitive Diagnosis and Its Implications for Testing. September 1987. Technical Report.
INSTITUTION Pittsburgh Univ., Pa. Learning Research and Development Center.
SPONS AGENCY Office of Naval Research, Arlington, Va.; Swedish Council for Research in the Humanities and Social Sciences, Stockholm.
REPORT NO UPITT-LRDC-ONR-KUL-8702
PUB DATE 88
CONTRACT ONR-N00014-85-K-0337
NOTE 77p.; In: Frederiksen, Norman; Glaser, R.; Lesgold, A. M.; Shafto, M., Eds. Diagnostic Monitoring of Skill and Knowledge Acquisition. Erlbaum, Hillsdale, NJ.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Adults; *Cognitive Measurement; Cognitive Psychology; College Students; *Computer Simulation; Higher Education; Item Analysis; *Problem Solving; *Psychometrics; *Spatial Ability; *Standardized Tests; Test Construction
IDENTIFIERS Reasoning Tests; Spatial Tests; *Trace Analysis

ABSTRACT

Recent theoretical developments in cognitive psychology imply both a need and a possibility for methodological development. In particular, the theory of problem solving proposed by Allen Newell and Herbert A. Simon (1972) provides the rationale for a new empirical method for the processing of think-aloud protocols--trace analysis. A detailed example is presented in which trace analysis is applied to human performance on a spatial reasoning task. The study, a portion of a larger study, involved 12 subjects who were required to solve a variety of spatial arrangement problems under different conditions. The protocol produced by a single subject--a 30-year-old female psychology student--while solving the block problems was selected for analysis on the basis of completeness and interest. The relations between trace analysis, on the one hand, and the psychometric ideas of measurement and standardization, on the other, are discussed. A non-psychometric approach to standardized testing, referred to as "theory referenced test construction," is proposed. The main premise of the approach is that test items should be validated against computer-implemented information processing models of the relevant cognitive functions. Eight figures and a 64-item list of references are provided. A computer simulation program, a bibliography of Knowledge and Understanding in Human Learning reports, and a report distribution list are appended. (TJH)

Trace Analysis and Spatial Reasoning:
An Example of Intensive Cognitive Diagnosis and its Implications for Testing

Stellan Ohlsson
Learning Research and Development Center,
University of Pittsburgh
Pittsburgh, Pennsylvania 15260

Technical Report No. KUL-87-02

September, 1987

LEARNING RESEARCH AND DEVELOPMENT CENTER

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. AUG

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

University of Pittsburgh

BEST COPY AVAILABLE

Trace Analysis and Spatial Reasoning:

An Example of Intensive Cognitive Diagnosis

and its Implications for Testing

Stellan Ohlsson
Learning Research and Development Center,
University of Pittsburgh
Pittsburgh, Pennsylvania 15260

Technical Report No. KUL-87-02

September, 1987

To appear in N. Frederiksen, R. Glaser, A. M. Lesgold, and M. Shafto, (Eds.), *Diagnostic monitoring of skill and knowledge acquisition.* Hillsdale, NJ: Erlbaum.

Reproduction in whole or in part is permitted for any purpose of the United States Government. Approved for public release; distribution unlimited.

Copyright © 1988 Stellan Ohlsson

The preparation of this manuscript was supported in part by ONR contract N00014-85-K-0337 and in part by a grant from the Swedish Council for Research in the Humanities and Social Sciences. The opinions expressed do not necessarily reflect the position of the sponsoring agencies, and no endorsement should be inferred. Goran Hagert was a constant and invaluable discussion partner during the early phase of the research, and Chris Wolfe played a similar role during the final phase. I thank Susanne Lajoie for helpful comments on an earlier version of the chapter, and editor Norman Frederiksen for his great patience with my extensive reworkings of the original draft.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release, distribution unlimited
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) UPITT/LRDC/ONR/KUL-8702	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Learning Research & Development Center, University of Pittsburgh 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Cognitive Science Program Office of Naval Research (code 1142CS)
6c. ADDRESS (City, State, and ZIP Code) 3939 O'Hara Street Pittsburgh, PA 15260	7b. ADDRESS (City, State, and ZIP Code) 800 North Quincy Street Arlington, VA 22217-5000
8a. NAME OF FUNDING / SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-85-K-0337
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61153N; PROJECT NO. RR04206; TASK NO. RR040206-0A; WORK UNIT ACCESSION NO NR442a-545

11. TITLE (Include Security Classification)
TRACE ANALYSIS AND SPATIAL REASONING: AN EXAMPLE OF INTENSIVE COGNITIVE DIAGNOSIS AND ITS IMPLICATIONS FOR TESTING (UNCLASSIFIED)

12. PERSONAL AUTHOR(S)
Stellan Ohlsson

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM _____ TO _____	1987, September	

16. SUPPLEMENTARY NOTATION
To appear in N. Frederiksen, R. Glaser, A.M. Lesgold, and M. Shafto, (Eds.), Diagnostic monitoring of skill and knowledge acquisition. Hillsdale, NJ: Erlbaum.

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP
05	10	
23	02	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Cognitive diagnosis; protocol analysis; spatial reasoning; standardized testing

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Recent theoretical developments in cognitive psychology imply both a need and a possibility for methodological development. In particular, the theory of problem solving proposed by Allen Newell and Herbert A. Simon provides the rationale for a new empirical method that here will be called "trace analysis." A detailed example is presented in which trace analysis is applied to human performance on a spatial reasoning task. The relations between trace analysis, on the one hand, and the psychometric ideas of measurement and standardization, on the other, are discussed. A non-psychometric approach to standardized testing, called "theory referenced test construction," is proposed. The main idea of theory referenced test construction is that test items should be validated against computer-implemented information processing models of the relevant cognitive functions.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	Unclassified
22a NAME OF RESPONSIBLE INDIVIDUAL Susan M. Chipman	22b. TELEPHONE (Include Area Code) 202-696-4318
	22c. OFFICE SYMBOL ONR 1142pt

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

Knowledge and Understanding in Human Learning

Knowledge and Understanding in Human Learning (KUL) is an umbrella term for a loosely connected set of activities lead by Stellan Ohlsson at the Learning Research and Development Center, University of Pittsburgh. The aim of KUL is to clarify the role of *world knowledge* in human thinking, reasoning, and problem solving. World knowledge consists of general principles, and contrasts with facts (episodic knowledge) and with cognitive skills (procedural knowledge). The long-term goal is to answer four questions: How are new principles acquired? How are principles utilized in insightful performance? How are principles utilized in learning to perform? How can instruction facilitate the acquisition and utilization of principled (as opposed to episodic or procedural) knowledge? Different methodologies are used to investigate these questions: Psychological experiments, computer simulation, historical studies, semantic, logical, and mathematical analyses, instructional intervention studies, etc. A list of KUL reports appear at the back of this report.

Table of Contents

Abstract	3
1. On Methodology	4
2. The Enaction Theory and Trace Analysis	8
2.1. The Enaction Theory of Thinking	9
2.2. The Method of Trace Analysis	11
3. Trace Analysis Applied to Spatial Reasoning	14
3.1. The Subject and the Behavioral Record	17
3.2. Diagnosing the Subject's Problem Space	19
3.3. Diagnosing the Subject's Solution Path	25
3.4. Diagnosing the Subject's Strategy	32
3.5. Evaluation the Strategy Hypothesis	37
3.6. A Do-It-Yourself Summary	40
4. Implications for Standardized Testing	41
4.1. Trace Analysis and Measurement	42
4.2. Trace Analysis and Standardization	43
4.3. Towards Theory Referenced Test Construction	45
References	50
Appendix A. Simulation program for SI6	50
KUL reports	56

Abstract

Recent theoretical developments in cognitive psychology imply both a need and a possibility for methodological development. In particular, the theory of problem solving proposed by Allen Newell and Herbert A. Simon provides the rationale for a new empirical method that here will be called *trace analysis*. A detailed example is presented in which trace analysis is applied to human performance on a spatial reasoning task. The relations between trace analysis, on the one hand, and the psychometric ideas of measurement and standardization, on the other, are discussed. A non-psychometric approach to standardized testing, called *theory referenced test construction*, is proposed. The main idea of theory referenced test construction is that test items should be validated against computer-implemented information processing models of the relevant cognitive functions.

1. On Methodology

Mental life is invisible and its expression in action is under voluntary, intentional control. The psychological sciences have been slow in accepting the methodological challenge posed by these two facts. Several evasive tactics have been tried. The first tactic was to observe mental life directly, by looking inward. The second evasive tactic was to decree that action itself is the object of study in psychology. Both of these tactics deny the necessity of inferring mental events from observations of actions. In a third evasive move psychology was declared a part of the humanities, with the implication that interpretation of human behavior is necessarily, irrevocably subjective. While admitting the need for inferences this stance denies the possibility of imposing a discipline on those inferences, a discipline which makes rational discussion and intersubjective agreement possible. We now know that the evasive tactics of introspectionism, behaviorism, and humanistic psychology do not work; they were worth trying, but they failed. We are left with the sole option of tackling the methodological challenge of mental life head on.

One might take the view that a scientist should attack significant substantive problems, propose interesting theories, and discover novel facts. If he[1] does, the methodological development of his science will take care of itself. Methodology *per se* is boring, unending fiddling with technicalities, an activity best left to the pedantic introvert who brings no creativity to his work. A real scientist worries about ideas and problems, not about methods.

There are several mistakes hiding in this proud attitude. First, careful observation of scientific research by a knowledgeable and sympathetic observer like Toulmin (1972) has revealed that the knowledge transmitted by one generation of scientists to the next does not consist mainly of particular explanations, but, instead, of the procedures by which explanations are constructed. There is, then, evidence that our methods are closer to the center of scientific knowledge than the traditional disdain for methodological work admits. Second, methodology has to be distinguished from the perfecting of measuring instruments. Methodology certainly deals with the accuracy of observations in general and the precision of measurements in particular. But the core topics of methodology are: the nature of evidence, forms of description, patterns of inference, boundary conditions on the validity of inferences, the design of explanatory procedures, and the standards by which particular explanations are judged. Third, scientific

[1]For convenience I am using "he", "his", etc. to refer to both genders throughout the chapter.

breakthroughs are brought about by new methodologies as well as by new ideas. One need only mention the electron microcope, the carbon-14 dating method, and the cyclotron. Fourth, in the application of science to practical concerns methods are often useful even in the absence of theory. For instance, the method of ascertaining verticality by suspending a weight from a string is useful for building a house even in the absence of a theory of gravitation. Methodology is essential both for the creation and the application of scientific theories.

Methodological innovation has not been a conspicuous feature of psychological research. The evasive tactics mentioned above discouraged serious thinking about how to infer states of mind from observations of behavior. Methodological development was restricted to the design of new statistical procedures, and methodological knowledge became limited to knowledge about the proper application of such procedures. But the cognitive revolution (Gardner, 1985) puts methodological innovation on the psychologist's agenda. Cognitive psychologists are collecting new types of data in support of new types of theories. We need a new view of methodology, new concepts to replace the stale dichotomies that dominated methodological debate in the past (description vs. hypothesis testing, experimental vs. correlational, laboratory control vs. ecological validity, objective vs. subjective, research vs. application, standardized vs. clinical, etc.).

In order to take a fresh look at the fundamental dimensions of psychological methods, consider the following formulation of the basic problem: *Given a behavioral record of person P (at time t), infer a description of P's mental state (at t).* This formulation implies that the three fundamental dimensions of psychological methods are (a) the type of *behavioral record* to which a method applies (i. e., the input), (b) the type of *description* of mental states that a method generates (i. e., the output), and (c) the rules of inference--or, in the terminology of Toulmin (1972)--the *explanatory procedures* that are used to construct the description, given the behavioral record (i. e., the transformation of the input into the output).

With respect to input, we can distinguish between extensive and intensive methods. *Extensive methods* rely on relatively shallow analysis of a large number of performances, while *intensive methods* rely on a deep analysis of a small number of performances (possibly just one; see Dukes, 1968). For instance, the methods used by experimental psychology and by psychometrics are extensive, while the methods used by psychoanalysts are intensive. Furthermore, behavioral records vary with respect to whether they preserve *sequential* information or not, and methods that do preserve sequential information vary with respect to the *temporal density* of that information.

With respect to output, we can distinguish between singleton and aggregate descriptions. *Singleton descriptions* summarize observations that derive from a single individual, while *aggregate descriptions* summarize observations which derive from a group of individuals. For instance, psychometric and psychoanalytic methods produce singleton descriptons, while the methods used in experimental psychology typically produce aggregate descriptions.

With respect to explanatory procedures, we can distinguish between open and closed methods. The purpose of *open methods* is to reveal the structure in the behavioral record. Open methods proceed in bottom-up fashion from the data towards the description. The purpose of *closed methods* is to ascertain how closely the behavioral record fits a pre-defined structure. For instance, the methods used in psychoanalysis are typically open methods, while the methods used in experimental psychology are closed methods. The psychometric tradition has a double-sided relation to this dimension. The *construction* of tests use open methods like factor analysis and cluster analysis, but the *application* of a test battery, once constructed, is an instance of a closed method.

In summary, I suggest that psychological methods should be discussed in terms of the type of behavioral records they apply to, what type of descriptions of mental states they generate, and what type of explanatory procedures they use to transform the record into the description. The rest of this chapter presupposes this schema for the analysis of methods.

A major new type of behavioral record introduced into cognitive psychology in recent years[2] is that of *protocols*, in particular think-aloud protocols (Newell, 1966; Newell & Simon, 1972; Ericsson & Simon, 1984; Williams & Hollan, 1981; Williams & Santos-Williams, 1980). A protocol is a verbatim transcript of spontaneous talk on the part of a subject about a task. There are two frequently used methods for the processing of protocols. The simplest is the *method of excerpts* which has been practiced in the humanities for a long time. It consists in selecting a part of the corpus and printing it in full, thus letting the reader see for himself, as it were. The excerpt is selected so as to exhibit a typical case, to prove the existence of some phenomenon, or to make a point of some kind; frequently, two excerpts are shown side by side in order to illustrate a difference or a contrast.

The other popular method for processing verbal protocols is known in social psychology as *content*

[2]The use of verbal protocols was not *invented* in recent years, but rather re-discovered. See the historical section in Ericsson and Simon (1984, pp. 48-61).

analysis[3] (Holsti, 1968). In content analysis one proceeds by defining a set of categories of textual events and counting the frequency with which each category occurs in a corpus of protocols. These frequences can be used as dependent variables in experimental studies. Cognitive psychologists re-invented this method and have used it frequently in recent years, without, however, paying attention to the rather extensive experience of social psychologists with respect to its applicability, reliability, and validity (Holsti, 1968).

Newell (1966) and Newell and Simon (1972) have proposed a new method for the analysis of protocols. They did not name their method; for convenience, I will refer to it as *trace analysis*. In the terms introduced above, trace analysis is an intensive, open method which aims for singleton descriptions. The type of behavioral record to which trace analysis applies is a think-aloud protocol. The type of description produced is a specification of an information processing system that behaves like the observed person. The explanatory procedures that generate an information processing system from a think-aloud protocol are rather complicated; they will be presented below in the context of an example. Trace analysis breaks new ground in that it combines an interest in the meaning of protocol fragments (which is characteristic of the method of excerpts) with a concern for imposing a discipline on the process of analysis (which is characteristic of content analysis). Also, it makes use of the sequential information in a protocol, a type of information which is destroyed by methods that build on category frequency.

Trace analysis has been all but ignored. Today, sixteen years after its introduction, there exists, to the best of my knowledge, no published research report that uses it, other than the book in which it was originally introduced. One possible explanation for this fact is that the description of the method is somewhat obscure, and, moreover, buried in a single chapter of a large and rather difficult book (Newell & Simon, 1972, Chap. 6). Another possible explanation is that Newell and Simon introduced trace analysis in the context of a specific application, namely a study of so-called cryptarithmetic problems.[4] Since human performance on cryptarithmetic problems is not a hot *substantive* topic researchers might bypass Newell and Simon's study as not relevant to their interests, thus missing the *methodological* contribution of that study. Also, researchers might fail to distinguish between different *types* of protocol analysis.

---

[3]This is an unfortunate misnomer. For content analysis to yield intersubjectively valid results, the categories used must be defined on the basis of syntactic, lexical, or other criteria which ignore content.

[4]In cryptarithmetic problems words are treated as numbers, as in SEND + MORE = MONEY. The task is to replace the letters with digits in such a way that the arithmetic operation is correct.

Researchers who use either the method of excerpts or the method of content analysis may believe that they *are* using the method proposed by Newell and Simon, and consequently feel no need to study their original description of trace analysis. Yet another possible explanation is that trace analysis breaks so radically with the methodological traditions of academic psychology that it simply has not been understood.

The purpose of this chapter is to develop the implications of trace analysis for standardized testing, and to facilitate and promote wider discussion and use of trace analysis in both research and practical contexts. The introduction to trace analysis presented here is, I believe, more accessible than the original presentation by the inventors of the method. Also, the task domain chosen for the application--verbally presented spatial reasoning problems--is different enough from cryptarithmetic to provide some evidence for the generality of the method.

The chapter is organized as follows. Section 2 puts forth the rationale of trace analysis. Section 3 is devoted to an application of trace analysis to spatial reasoning. Section 4 contains a speculative proposal for a non-psychometric methodology of standardized testing that builds on trace analysis.

2. The Enaction Theory and Trace Analysis

Allen Newell and Herbert A. Simon have proposed that we think by mentally enacting alternative sequences of actions with respect to a problem (Newell, 1966, 1980, 1987; Newell, Shaw, & Simon, 1958; Newell & Simon, 1972). Although they did not name their theory, I have called it the *Enaction Theory* in other contexts (Ohlsson, 1983) and I will continue to do so here. The main methodological implications of the Enaction Theory are that cognitive diagnosis should be based on a sequentially ordered and temporally dense trace of the performance to be diagnosed, and that a diagnostic description should take the form of a specification of an information processing mechanism that can reproduce the observed performance. Think-aloud protocols fulfill the methodological requirements better than other types of behavioral records. Trace analysis is primarily a method for the analysis of think-aloud protocols. Both the Enaction Theory and the method of trace analysis are described below.

2.1. The Enaction Theory of Thinking

The Enaction Theory asserts that cognitive processing takes the form of *heuristic search through a problem space*. The process of heuristic search consists in using a *strategy*, i. e., a collection of problem solving heuristics, in order to decide which *operator*, i. e., cognitive skill, should be applied to to the current *knowledge state*, i. e., mental representation of a problem. The application of an operator generates a new knowledge-state. The successive application of operators continues until a knowledge-state is reached in which the problem solver's *goal* is satisfied. These concepts may need some clarification.

Consider a person confronted with an intellectual task, such as the Tower of Hanoi puzzle, a chess problem, an algebra problem, Maier's Two-String Problem, or a geometric proof problem. In order to solve the task he must construct a mental representation of the given information, the problem-as-presented. The internal description of the problem is called the *initial knowledge state*. For instance, in the Tower of Hanoi puzzle[5] the problem-as-presented can be seen as *a pyramid of discs*; in a verbal reasoning task the givens might be conceptualized as *a list of related facts*. The problem solver must also build a mental representation of what he is supposed to do with the task, i. e., of what counts as having solved it. This representation is his *goal*. The goal specifies when to terminate the problem solving effort. For instance, in the Tower of Hanoi puzzle the goal might be conceptualized as *transport the pyramid of discs to another peg*. The initial knowledge state and the goal together constitute an understanding of the problem.

Once the task has been understood, the thinker must call up a repertory of mental actions or cognitive skills with which he can process the problem. They are called *operators*, because they operate upon the current mental representation of the problem to generate a new representation (namely a representation of what the problem situation would be like if the physical action corresponding to the operator were to be carried out). The application of operators is a mental, rather than a behavioral, process. The theory asserts that the thinker is acting out *in his mind* what would happen if such and such an action were to be taken with respect to the problem. For instance, in solving a chess problem the thinker is likely to imagine what would happen, if he were to make such and such a move; in an algebraic proof problem, the thinker might anticipate what a particular formula would look like, if a certain transformation were applied to it.

---

[5]Given three pegs and *N* discs of different sizes stacked on one of the pegs in order of increasing size, move the discs to another peg by moving one disc at a time, without ever putting a larger disc on a smaller (Simon, 1975).

The theory claims that the problem solver at any one time considers only a small ensemble of operators that he has judged as relevant for his current problem. The problem solver may or may not be correct in his relevance judgments, so the operator ensemble may or may not include all operators necessary to solve the problem.[6]

The initial knowledge state and the repertory of relevant operators (or the operators the problem solver believes are relevant) implicitly specify a space of solution candidates to the problem, known as the *problem space*.[7] A solution consists in the application of some operator to the initial state, then another (not necessarily distinct) operator to the resulting state, then yet another operator to its result, etc., until the goal has been reached. A solution candidate consists in a sequence of operator applications, known as a *path* through the problem space. For instance, *pick up the hammer, tie the hammer to one of the ropes, set the rope swinging, walk over to the other rope, grab the first rope as it comes swinging, untie the hammer, and tie the ropes together* is a sequence of steps which constitutes a solution to Maier's Two-String Problem.[8] The initial state and the repertory of operators together generatively define the set of all possible solution candidates. The Enaction Theory asserts that thinking consists in the mental exploration of this set.

In routine action the sequence of operators that lead to the goal is known beforehand. For instance, in solving a multi-column addition task, any competent adult knows to begin with the column to the right, add within a column, carry to the next column to the left, etc. Such a task is not properly called a *problem*. A task is a problem when the solution path is not known beforehand, but has to be found by trying out various operator sequences, judging how promising they are, and selecting one for execution. If the selected action sequence does not, in fact, lead towards the goal, the problem solver has to go back and try a different sequence, a process that naturally enough is called *back-up*. The process of exploring alternative paths is called *search*. The search is anticipatory; we search in the head before we search in the flesh, as it were, a decision making technique that has considerable survival value.

A problem space can be searched systematically, by exploring all possible paths. But simple

---

[6]This principle has been used to explain the phenomena of restructuring and insight in problem solving (Ohlsson, 1984, c).

[7]The terminology chosen by Newell and Simon is unfortunate on this point. "Solution space" would have been more descriptive than "problem space". Grave misunderstanding of the theory results if a problem space is construed as *a space of problems* instead of as *a space of solution candidates for a particular problem*.

[8]Two ropes are suspended from the ceiling; the distance between them is too wide to allow a person to reach one rope while holding the other. A variety of everyday objects is provided. The task is to tie together the two ropes (Maier, 1970).

combinatorial calculations will show that the number of possible operator sequences is astronomical, even if the repertory of actions is small and the length of the solution path short. For instance, if there are 5 relevant operators and if the solution path is 10 steps long, then there are 5 to the 10th power, or approximately ten million, different solution candidates. Systematic search is not feasible. Instead, the Enaction Theory claims, problem solvers search selectively, applying rules of thumb called *heuristics*. Such a rule contains information about which operator is most likely to lead towards the goal in some particular type of situation. For instance, a useful heuristic for geometry proof problems is *if the task is to prove two geometric objects congruent, and if the given figure contains many straight lines, try to find congruent triangles*. A problem solving strategy consists of a collection of such rules. The efficiency of problem solving is a function of how accurately the available heuristics sort out blind alleys and focus the search on a path that leads to the goal. The Enaction Theory explains expert performance in knowledge-rich domains (Newell & Simon, 1972, Chap. 11-13) as a product of a large number of very selective heuristics.

The Enaction Theory is a successful theory. The notion of heuristic search through a problem space has now been articulated with respect to a wide range of human behaviors, from syllogistic reasoning (Newell, 1980) to the configuration of computers (Rosenbloom, Laird, McDermott, Newell, & Orciuch, 1985). The theory explains why some problems are more difficult than others (see, e. g., Kotovsky, Hays, & Simon, 1985). It explains individual differences in thinking (see, e. g., Newell & Simon, 1972, Chaps. 7, 10, and 13). During recent years the Enaction Theory has been the basis for several theories of learning (see the collections of articles edited by Anderson, 1981; by Bolc, 1987; and by Klahr, Langley, & Neches, 1987a). The Enaction Theory carries definite implications for education (Frederiksen, 1984; Ohlsson, 1983; in press); indeed, it is solid enough to support the design of intelligent tutoring systems (Anderson, Boyle, & Reiser, 1985). There is at the current time no other theory of human thinking with comparable scope, precision, empirical grounding, and practical utility.

2.2. The Method of Trace Analysis

If the Enaction Theory of thinking is correct, what kind of empirical method do we need in order to explain particular problem solving performances? The theory implies that a psychological explanation consists of three parts: An hypothesis about the subject's problem space (his understanding of the problem, and the mental resources he has available for processing it), an hypothesis about his solution path (the sequence of mental states he traversed on his way to the goal), and an hypothesis about his

strategy (the collection of heuristics that generated the solution path). The empirical observations we collect and the procedures by which we analyze them must enable us to identify those three constructs.

Newell and Simon (1972) proposed that think-aloud protocols is an ideal type of behavioral record for the study of problem solving, and they invented *trace analysis*[9] as a method for the processing of such protocols. The main methodological works on trace analysis are Newell and Simon (1972, Chapter 6) and Ericsson and Simon (1984). Trace analysis proceeds in a bottom-up fashion through three main steps:

1. Construct the subject's problem space: (a) infer his *mental representation* of the task from the words he uses to describe the problem; (b) infer his *ensemble of operators* from recurring patterns of activity that give rise to new conclusions; and (c) infer his *goal* by noticing when, under what conditions, he declares himself finished with the task.

2. Identify the subject's solution path by making use of the sequential information in the protocol in order to map it onto the problem space identified in step 1. This amounts to choosing a path through the problem space which explains as many of the events in the protocol as possible.

3. Hypothesize the subject's strategy by inventing problem solving heuristics that can reproduce his solution path. The strategy hypothesis is complete if for each state-step pair along the solution path, there is some heuristic in the strategy that can generate that step when applied in that state.

The description of the *subject* achieved with this method consists of a problem space and a strategy for how to search that space. The description of his *performance* consists of a solution path.

The three steps described above build on each other: Identification of the problem space enables the description of the solution path, and a description of the solution path enables identification of the heuristics. Only the first two steps build directly on the information in the data. The step of identifying the problem space makes use of the content of the protocol utterances, while the step of laying out the solution path builds on the sequential information in the protocol. The third step, however, builds *on the previous two steps.* The problem solving heuristics used by the subject are inferred from the solution path, not from the protocol. In summary, the problem space constitutes a special-purpose formalism for describing the solution path; the solution path is a low level mini-theory which explains the behavioral

[9]The name "trace analysis" is preferred over "protocol analysis", since I do not want to imply that the method invented by Newell and Simon is the *only* possible method for the analysis of protocols.

record; the strategy is slightly-higher-level mini-theory that explains the solution path.[10]

The Enaction Theory implies two methodological requirements that are difficult to fulfill with any other type of behavioral records than think-aloud protocols. The first requirement is that the behavioral record must enable us to infer the subject's conceptualization of the problem. We therefore need to hear him talk about the problem. How does he parse the problem situation into distinct objects, what properties does he assign to them, and what relations does he see between them? What representational formats does he use to encode those properties and relations? For instance, in so-called cryptarithmetic problems, the concept of *parity*--whether a number is odd or even--is often crucial to successful problem solving (Newell & Simon, 1972). It is obviously difficult to know whether a person is using the concept of parity or not, unless we hear him talk about the problem. As a second example, Johnson-Laird (1983) has argued that people solve verbal reasoning problems with mental models, rather than with propositional representations. It is obviously difficult to know what representational format a person is using, unless we hear him verbalize it.

The second methodological requirement of the Enaction Theory is that the behavioral record must enable us to infer the *sequence* of mental events that took place when the subject solved the experimental problem. Unless we know the solution path, we cannot infer the strategy. Different paths might lead to the same end-state, so a recording of the end-state or the time it took the subject to arrive at the end-state does not enable us to identify his path. We need to observe the intermediate stages of the problem solving effort, the sequence of partial results created along the path to solution. The trace of the partial results should preferably be *temporally dense*, i. e., have many observations of the performance per unit of time, in order to accurately discriminate the subject's path from alternative paths through the problem space.

Think-aloud protocols fulfill both of the above requirements. They reveal how subjects conceptualize the experimental problem, and they provide a sequentially ordered and temporally dense trace. Other types of behavioral records are less satisfactory. Interviews destroy sequential information, because the order of the subject's utterances is partially controlled by the order of the interviewer's questions. In retrospective interviews the sequential information is further corrupted by memory failures. In general,

[10]The hierarchy of explanations does not end with the strategy, of course. The strategy is explained by a learning theory, which, in turn, is explained by the structure of the cognitive architecture; the latter is related to the structure of the brain; and so on.

interviews reveal the subjects's representation, but does not enable us to infer his solution path. Video tapes of behavior or the recording of key strokes on computer terminals provide sequential information, but they do not give us any insights into the subject's mental representation. In general, behavioral recordings reveal the path, but not the representation. Eye movement recordings may reveal the representation (since they tell us what features of the problem situation the subject attends to, or can discriminate between), but since they do not reveal what the subject *does* with the problem information, they do not enable us to infer the solution path. In short, think-aloud protocols fulfill the methodological requirements of the Enaction Theory better than other types of behavioral records.

In summary, human beings are hypothesized to think by mentally exploring alternative paths through some search space. The methodological implications of this hypothesis is that cognitive diagnosis should be based on a sequentially ordered and temporally dense behavioral record that is analyzed with the goal of designing an information processing mechanism that can reproduce the observed behavior. A concrete example of this kind of cognitive diagnosis is worked out in detail in the next section. The implications of this methodology for the construction of standardized tests are developed in the fourth and final section.

3. Trace Analysis Applied to Spatial Reasoning

Consider the spatial reasoning problems in Figure 3-1. Each problem consists of a short text describing a static situation by asserting certain spatial relations between some discreet, stable objects. It ends with a question concerning a relation not explicitly mentioned in the text. I call problems of this sort *spatial arrangement problems*. The relational concepts used are common sense spatial concepts.[11] They include unary predicates like "bottommost", tertiary predicates like "between", and ambiguous predicates like "adjacent". If the number of objects in such a problem is larger than three, it will usually take an adult more than a minute to solve that problem; if the number of objects is, say, ten, and if the relational structure embedded in the premises is complex, the solution time can be as long as 20 minutes.

From a problem solving point of view, spatial arrangement problems are unusual in that they are static. Many problems used to study problem solving require a sequence of transformations of the given situation. In a spatial arrangement problem, on the other hand, the task is not to transform the given

[11]The problem texts are translated from Swedish. Phrases like "bottom-most but one" and "frontmost" may not be good English, but their Swedish counterparts are quite idiomatic.

1. The Bench Problem

Some boys are sitting on a bench.

Jonas is further right than Ingvar.
Olof is further left than Ingvar.
David is immediately to the left of Jonas.

Who is immediately to the right of Ingvar?

2. The Block Problem

A child is putting blocks of different colors on top of each other.

A black block is between a red and a green block.
A yellow block is further up than the red one.
A green block is bottommost but one.
A blue block is immediately below the yellow one.
A white block is further down than the black one.

Which block is immediately below the blue one?

3. The Ice-Cream Problem

Some boys are standing in line at an ice-cream stand.

Rolf is further towards the front than Erik.
Sven is further towards the front than Ove.
Nils is immediately behind Mats.
Hans is frontmost but one.
Mats is further back than Ove.
Erik is immediately behind Hans.
Leif is further back than Mats.

Who is immediately behind Erik?

Figure 3-1: Examples of spatial arrangement problems.

situation, but to understand it well enough to answer a question. From a psychometric point of view, spatial arrangement problems would be expected to have high loads on spatial ability, reasoning ability, and verbal ability. A main difference between spatial arrangement problems and typical test items is that spatial arrangement problems take more time to solve.

Empirical studies of spatial arrangement problems, using both trace analysis and experimental methods, have revealed a number of phenomena:

- A majority of adults solve spatial arrangement problems with the help of a mental model,[12] rather than by reasoning exclusively in a propositional mode (Hagert, 1980a, 1980b; Johnson-Laird, 1983; Ohlsson, 1980a, 1984b).

  A small minority of adults use a propositional reasoning method based on the idea of elimination of alternatives (Ohlsson, 1980a, 1984b).

  An even smaller minority try to apply other, less rational approaches to the problem, such as trying to infer the quantitative distances between the objects (Ohlsson, 1980a).

- The particular problem spaces used to implement the mental model building strategy vary from one individual to the next, as do the heuristics used to search them, with substantial differences in the solution paths traversed by different persons as a consequence (Ohlsson, 1980a, 1980b, I982).

- Some subjects shift back and forth between model-building and propositional strategies. Subjects can be induced to make such strategy shifts, even when they do not show any spontaneous tendency to do so (Ohlsson, 1984a).

- Strategies for spatial arrangement problems have a large attention allocation component. The solution to a spatial arrangement problem depends crucially upon which premises are read in which order. Consequently, differences in attentional heuristics is a major source of individual differences in this task domain (Ohlsson, 1984b).

- The spatial competence needed to solve spatial arrangement problems is large. A list of the inferences needed to build mental models of linear orderings from propositional descriptions contains over one hundred distinct inference patterns (Ohlsson, 1980a).

- Backups are frequent events in problem solving efforts in this domain. However, a large proportion of backups are not followed by the exploration of new search paths, but by the re-traversal of the already explored search path (Hagert & Rollenhagen, 1981; Ohlsson,

[12]The term "mental model" is here used in the sense of Johnson-Laird (1983), who defines a model as an object which satisfies a set of propositions. This is the sense in which the term is used in the study of formal logic. The term is commonly used within cognitive science to refer to any integrated knowledge unit with a large grain size, particularly if it encodes knowledge about a physical mechanism or process. For examples of this alternative use of the term, see the collection of articles by Gentner and Stevens (1983).

1980a). Hence, they are not backups in the search theory sense. These backups occur, I believe, because working memory capacity limitations make it necessary to recreate intermediate results from time to time. I will call backups which are followed by repetition of previously performed inferences *consolidation backups*.

The general conclusions summarized above are based on large numbers of applications of trace analysis. For example, Study II of Ohlsson (1980a) was based on fifty protocols, each of which was analyzed with the help of trace analysis. A detailed diagnosis of a single performance will be presented in detail.

3.1. The Subject and the Behavioral Record

The performance to be diagnosed here was selected from a larger study (Ohlsson, 1980a, Study I). Twelve subjects participated in the study. They solved a variety of spatial arrangement problems under different conditions. The protocol to be discussed here was produced by a subject labeled SI6 while solving the Block Problem (see Figure 3-1). It was selected for analysis on the basis of completeness and interest.

Subject SI6 was a 30 year old psychology student. She participated in the experiment as part of a course requirement. She was not paid. The Block Problem was her third problem in the experimental session. In a previous session she had solved three simpler spatial arrangement problems.

The problem text was typed as it appears in Figure 3-1 on a white index card which was handed over to the subject at the beginning of the solution attempt. She had the card available throughout the solution attempt. She was not allowed the use of paper and pencil or any other tool. She was instructed to think aloud. The exact instruction given was "give words to your thoughts as you have them". She was instructed to begin her solution attempt by reading through the problem text aloud. The verbalizations were tape recorded and transcribed verbatim.

The complete protocol[13] is shown in Figures 3-2 and 3-3. F-numbers in the following analysis refer to protocol fragments in those figures. The protocol is 3:40 minutes long (220 seconds), including the initial reading of the problem text. It contains a total of 314 words, which means that the subject's speech rate was approximately 1.4 words per second. There are no task-irrelevant passages in the protocol, nor any

[13]The subject spoke Swedish, so the text in Figures 3-2 and 3-3 is a translation of the original protocol.

F1. a child puts blocks in different colors on top of each other
F2. a black block is between a red and a green block
F3. a yellow block is further up than the red one
F4. a green block is bottommost but one
F5. a blue block is immediately below the yellow one
F6. a white block is further down than the black one
F7. what block is immediately below the blue one
F8. the black block is between a red and a green
F9. block
F10. well that does not mean that it must be exactly between
F11. could be something else between also
F12. a yellow block is further up than the red one
F13. a green block is bottommost but one
F14. a blue block is immediately below the yellow one
F15. the yellow one is higher up than the red one
F16. and immediately below the yellow one comes the blue one
F17. then comes a red one
F18. I'd say
F19. well
F20. a
F21. a yellow block is higher up than the red one
F22. a green block is bottommost but one
F23. a blue block is immediately below the yellow one
F24. below the yellow one is a blue block
F25. and a yellow block is higher up than the red
F26. below the yellow is then also a red
F27. a blue and a red are below the yellow one
F28. and a
F29. a blue and a red are under the yellow block
F30. and a green block is bottommost but one
F31. a black block is between the red and the green
F32. a black block
F33. a black block is between the red and the green block
F34. a white block is further down than the black one
F35. then there is a white
F36. and then we have a
F37. oh how difficult
F38. a white block is further down than the black one
F39. and the black one is between the red and the green

Figure 3-2: Think-aloud protocol from SI6 on the Block Problem, Part 1

F40. white red black green
F41. I'll say then
F42. but the green is bottommost but one
F43. then I'll say
F44. white green black red
F45. instead
F46. then the white one is bottommost
F47. white green black and red
F48. and then we had the
F49. blue one which is immediately below the yellow
F50. the yellow is higher up than the read
F51. then it is topmost so far
F52. the yellow one
F53. and the blue one is immediately below
F54. then it comes topmost but one
F55. which one is then immediately below the blue one
F56. immediately below the blue one is then the red one

Figure 3-3: Think-aloud protocol from subject SI6 on the Block Problem, Part 2.

interactions with the experimenter. The solution attempt ended when the subject gave her answer, which was correct.

3.2. Diagnosing the Subject's Problem Space

The problem space used by the subject is discussed in four subsections, dealing with her representation, her operations, her goal, and her memory resources, respectively.

Representation

The protocol shows that, as one would expect, the subject is capable of reading and comprehending the sentences in the problem text, and of making use of the propositional information conveyed by them. However, there are several classes of propositional constructions which are *not* used by this subject in this protocol. First, there are no examples of *negated* sentences in the protocol. SI6 does not use expressions of the form "Object X is not ...", e. g., "The black block cannot be above ... ". Second, there is no evidence for the use of *quantifiers*. She does not use expressions of the form "All objects are ..." or "At least on object is ...". Third, she does not use any *sentential connectives* (even though she uses "and" to connect arguments within propositions). In particular, she does not use any *if-then* constructs, such as "consequently", "therefore", "it follows that", etc. In summary, simple predicate-argument constructions are sufficient to capture the subject's representation of propositional information about the task.

```
<knowledge-state>  ::= <knowledge-element> /
                       <knowledge-element> <knowledge-state>

<knowledge-element>  ::= <tag> <knowledge-element>  /
                         <proposition> / <question>  / <model>

<proposition>  ::= (<predicate> <object-sequence>)

<question>  ::= (<predicate> ? <object>)

<predicate>  ::= ABOVE / IMMEDIATELY-ABOVE /
                 UNDER / IMMEDIATELY-UNDER /
                 TOPMOST / TOPMOST-BUT-ONE /
                 BOTTOMMOST / BOTTOMMOST-BUT-ONE /
                 ADJACENT / BETWEEN / ANSWER

<model>  ::= (<end-anchor>.1 <element-sequence> <end-anchor>.2)

<end-anchor>  ::= TOP / BTM

<element-sequence>  ::= <element> / <element> <element-sequence>

<element> ::= <object> / <relation>

<object-sequence>  ::= <object> / <object> <object-sequence>

<object>  ::= red / black / white / green / yellow / blue

<relation>  ::= <followed-by> / <adjacent-to>

<followed-by>  ::= "blank space"

<adjacent-to>  ::= "colon"

<tag>  ::= old / new / unc / imp

<probe> ::= FIRSTPREM / SECPREM / THIRDPREM /
            FOURTHPREM / FIFTHPREM / NEXTPREM / QUESTION

<operator> ::= READ / TRNS / INT / ANSW
```

Figure 3-4: Mental representation of subject SI6 for the Block Problem.

There is evidence in this protocol (as well as in other protocols from this subject) that the propositional format is not the only one used by SI6. In three places (F40, F44, and F47) she verbalizes her knowledge of the problem situation through a list of object-names, e. g.:

F40. white red black green

I take this as evidence that she is building a mental model of the problem situation, trying to see in her mind's eye the six blocks standing on top of each other.

An mental model of a linear ordering can be represented as a list of object symbols. Two refinements are needed to accurately represent this subject, namely end-anchors and a distinction between "adjacent" and "followed-by". First, the subject reads out her mental model in different directions at different times during the solution attempt (from top to bottom in F15-F17, and from bottom to top in, e. g., F40). This implies that her representation contains some device which allows her to keep track of the direction of a model. I will assume that she does this with the help of *end-anchors*, i. e., symbols which label the top and the bottom of the ordering respectively. In the formal model these are represented by the arbitrary symbols TOP and BTM, respectively.

Second, the subject is able to infer from premise 2 ("A yellow block is further up than the red one") and premise 4 ("A blue block is immediately below the yellow one") that the red block is below the blue block (see fragments F15-F17). This conclusion does not follow unless a distinction is made between two different relations, namely "x is adjacent to y", which implies that there is no object between x and y, and "x is followed by y", which does not say anything about proximity. Hence, the subject's mental model must contain some device for distinguishing between these two relations. In the formal model "adjacent to" is symbolized by a hyphen, and "followed by" with a blank space. For instance, (TOP x-y BTM) means that y is below and adjacent to x, (TOP x y-BTM) means that y is somewhere below x, that there could be other objects between x and y, and that there are no objects below y.

It will be necessary to assume that the various kinds of knowledge elements used to represent the problem have different *modes*. These modes will be symbolized in the analysis with the help of indices or *tags*. I will assume that the subject can tag knowledge elements in four different ways:

new	a new result (i. e., an output from an operator);
old	information which has already been used as basis for an inference;

unc a result which is experienced by the subject as unclear;

imp a result which is impossible because it contradicts the given information.

The evidence for the "new" and "old" tags is indirect. It consists in the global observations that SI6 always works on newly produced information, and that old information never confuses her or interferes with her processing. The evidence for the "unclear" status is more direct: In fragment F18 (see Figure 3-2) the subject directly verbalizes uncertainty about an outcome. The evidence for the "imp" tag, finally, is also direct: In the course of solving the problem the subject discovers a contradiction which leads her to revise her model; the fragments F42-F45 show that she is aware of this contradiction.

There are some types of information which are *not* used by SI6 in her solution to the Block Problem. First, she does not think about the *absolute positions* of the objects, in contrast to the relative positions the objects acquire in a partially completed model. For example, she does not ask herself questions like "What object goes into the topmost position?" or "What position should be assigned to object so-and-so?". Her representation is relative and topological in character, rather than absolute and positional.

A second and related point is that SI6 makes no use of *numerical information*. There is no evidence that she thinks in terms of number of objects: how many objects there are all in all, how many objects she has left to place, how many objects there could be room for in such-and-such a part of the model, etc. Indeed, there is no evidence that she ever counts the total number of objects mentioned in the problem. (This raises the question of how she knows that she has completed her mental model.)

Third, there are no verbalizations of *goals*, plans, or intentions. She never says anything about what she is trying to do, or what she would like to be able to do. e. g., "Next, I should find out the position of object X" or "I now want to find the object that is adjacent to object X".[14]

The representational format used by this subject on this type of task is summarized in a generative grammar on BNF form[15] in Figure 3-4.

Operators

[14]Other subjects in this study used position and numerical information in solving spatial arrangement problems, and gave clear evidence of setting themselves goals.

[15]The rules for the BNF notation can be found in many standard textbooks in computer science, and also in Newell and Simon (1972, pp. 44-46).

The subject shows evidence of using four basic problem solving operators (mental processes which produce new results): *reading* the problem text (READ), *translating* propositional information into a mental model (TRNS), extending an existing mental model by *integrating* further propositional information into it (INT), and *answering* a question by reading off the answer from a mental model (ANSW). They are defined in Figure 3-5.

It is worth emphasizing that the READ process is included among the problem solving operators. In this analysis reading new information from the display counts as a step forward in the problem space. This implies that a model of the subject's strategy must include assumptions about when and how she attends to the problem text. Heuristics for how to access the problem text play an important part in understanding human performance in this task domain.

The subject's world knowledge, or spatial competence, enters into the processing mainly through the TRNS, INT, and ANSW operators. They generate new conclusions. In order to model the subject's performance we need to know which spatial inferences these operators are capable of, i. e., what inferential competence we should stock them with, as it were, in order to accurately simulate human behavior. Task analysis indicates that there are approximately one hundred distinct inferences about linear orderings which adults in our culture would consider valid (Ohlsson, 1980a). The analysis of the inferential competence of this subject will not be pursued further here.

Goal

The goal of solving a spatial arrangement problem is to answer the question at the end of the problem text. It is trivial to answer questions about a linear ordering, if one has access to a complete model of that ordering, i. e., a model which includes all the objects mentioned in the problem text. I will assume that the operative goal of SI6 was to achieve a complete mental model. The evidence for this is that as long as her model is incomplete, she does not read the question she is supposed to answer. However, as soon as her model is complete in the sense of containing all the objects, she attends to the question and answers it.

How did the subject decide when her mental model was complete? Logically speaking, there are only two possibilities: to check that each object mentioned in the problem text is included in the model, or, alternatively, to count the objects in the model, count the objects mentioned in the text, and verify that the counts are the same. SI6 does not show evidence of carrying out either process. The protocol contains

READ(<probe>)	*Read from the problem text that item which is specified by the probe.* This operator accesses the external display, and delivers a proposition into working memory. The proposition is tagged as *new*, even if it has been read before. The probe is a description of that which is to be read. In the formal model, the probe can take the values FIRSTPREM, SECPREM, ... , etc., NEXTPREM, and QUESTION. These symbols are arbitrary, but their intended interpretation should be obvious.
TRNS(<proposition>)	*Translate a proposition into a mental model.* This operator takes a proposition as input, and delivers into working memory a model which satisfies that proposition. The proposition is tagged as *old* (given that the operator is successful), and the model as *new*. For instance, if the sentence "The blue block is immediately below the yellow one" (Premise 2 in the Blocks Problem) corresponds to the proposition "(Adjacent-Below blue yellow)", then TRNS[(Adjacent-Below blue yellow)] will result in the creation of the working memory element "(TOP yellow-blue BTM)".
INT(<proposition>)	*Integrate a proposition into the current model.* This operator takes a proposition as input, and tries to integrate its information into the current mental model. If it succeeds, it produces a new, extended model which is tagged as *new* and placed in working memory. The proposition is tagged as *old*. The previous model is deleted from working memory. For instance, if the current model is "(TOP yellow-blue BTM)", then INT[(Further-Below red yellow)] results in the extended model "(TOP yellow-blue red BTM)".
ANSW(<question>)	*Answer question.* This operator compares the question and the current mental model, and reads off the answer, if possible. The answer is then said, and the solution attempt ended. For instance, if the current model is "(TOP yellow-blue red BTM)", then ANSW[(Adjacent-Below blue ?)], where "(Adjacent-Below blue ?)" corresponds to the question "Which object is immediately below the blue block?", will result in the answer "red".

Figure 3-5: Basic problem solving operators of subject SI6.

no clue as to how she knows that her mental model is complete. Recognition of a complete model does not seem to be an explicit inferential step. I will assume that she infers that her model is complete when she fails to find any missing objects, i. e., any objects not yet included in the model. There is direct evidence for a process which tries to locate missing objects (see below).

Memory resources

A model of SI6's reasoning must make some assumptions about her working memory capacity and about her use of long-term memory. First, what working memory capacity should be presupposed? It turns out that a good account of the protocol can be constructed if we assume that this subject can reliably hold three knowledge elements in her head at any one time. (What counts as a knowledge element is defined by Figure 3-4.)

Second, the present analysis is based on the following hypotheses about long-term memory:

- The inferential knowledge needed to solve spatial arrangement problems is stored (procedurally) inside the TRNS, INT, and ANSW operators.
- Partial results are stored in long-term memory. More precisely, the current knowledge state is stored after each application of the TRNS and INT operators. Stored knowledge states can be retrieved and re-instated as the current state.[16]
- The long-term memory trace contains only the path from the initial state to the current state, i.e., search paths over which backups are made are deleted from memory.
- Long-term storage is used for various book-keeping purposes. For example, the READ operator is able to get the next premise from the problem text, i.e., the premise immediately below the last premise to be read. This presupposes some memory of which premise was last read. Similarly, the SCAN operator can continue a scanning pattern from the last point of scanning, which presupposes some memory of where the previous scan was broken off.

3.3. Diagnosing the Subject's Solution Path

Figures 3-4 and 3-5 define the subject's problem space. If the hypothesis they express is correct, they specify generatively the entire set of paths subject SI6 *could* have traversed while solving the Block Problem. The next step in the construction of an explanation of her performance is to identify which path she *actually* traversed. This is done by interpreting the protocol fragments in terms of the problem space

[16]Hence, a complete list of the subject's capabilities must include an operator that prepares for backup by storing the current state in long-term memory, and a backup operator which can retrieve a stored knowledge state. These operators are defined in Figure 3-7.

operators and their inputs and outputs. The following three interpretative principles were applied in the present analysis:[17]

1. Verbalizations from the subject are interpreted as outputs from operators, unless this would complicate the over-all interpretation.
2. Backups are assigned the shortest scope which is consistent with the evidence.
3. Verbalizations which are identical to sentences in the problem text are assumed to be the result of reading aloud from the problem card, unless this complicates the over-all interpretation. In cases of doubt, the audio tape was consulted.

These rules are applied below in mapping the protocol in Figures 3-2 and 3-3 onto the problem space defined by Figures 3-4 and 3-5. The result is an hypothesis about the subject's solution path that can be displayed graphically in the form of a so-called *Problem Behavior Graph* (PBG).[18] The PBG generated from the protocol in Figures 3-2 and 3-3 is shown in Figure 3-6. The first subsection below describes how the PBG is generated. The second subsection asks whether the path hypothesis reveals any unusual or special events, events which are in special need, as it were, of being explained.

Mapping the protocol onto the problem space

In the beginning the subject is simply reading the problem text, as she has been instructed to do (F1-F7). Presumably there is some change of goals between F7 and F8, from *read the text* to *solve the problem*, but there is no trace of it in the protocol. She then begins her solution attempt by reading the first premise (F8). Her next step cannot be interpreted within the problem space: She reflects on the meaning of the term "between" (F10-F11). This does not produce any new result in terms of the problem space. (This is the only step outside the problem space.) In F12 she is back in her attempt to solve the problem. She continues to read the premises in the order in which they are written, i.e., every time she reads, she reads the next premise (F12-F14). Upon reading the fourth premise, she notices the repeated occurrence of the yellow block, and begins to make inferences. The content as well as the phrasing of the fragments F16 and F17 implies knowledge of the internal relations between the yellow, blue, and red blocks. I interpret F16 as an application of the TRNS operator to premise four and F17 as an application of INT to premise two. The question of F15 then remains. The tone of voice on the tape does not

[17]The reader may want to compare the interpretative principles used here with the discussion of protocol interpretation in Ericsson and Simon (1984).

[18]The rules for PBG construction are set down by Newell and Simon (1972, p. 173). Briefly, time goes from left to right and from top to bottom. A knowledge state is a node, and an operator is a link. A backup results in a new node below the node representing the state to which the problem solver backed up; the two nodes are connected with a vertical line.

support the hypothesis that premise two is being re-read at this point. Working memory considerations show that premise two should still be available. Therefore, it has been interpreted as a rehearsal, i.e., not as a generation of a new result. The output from the sequence F15-F17 is probably tagged as "unclear" (F18), since it is followed by a consolidation backup (F19-F20).

The process is then repeated. In F21-F23 she reads again the premises in the order in which they are written on the problem card. There is one difference: after having translated premise four (F24), she re-reads premise two (F25) before it is integrated. No such re-reading was needed in the previous episode. However, as described above, in that episode she felt a need to rehearse premise two before translating premise four. She probably has some problem with working memory at this point, even though the assumption of a short-term capacity of three chunks predicts that premise two should still be available. At the end of this passage (F27-F29) she again has the result "yellow blue red".

She now continues by reading the two premises she has skipped, namely premise 3 (F30) and premise 1 (F31), and the premise she has not yet looked at, namely premise 5 (F34), in that order. In F32 she is trying to do something with the black block, but it is unclear what. She fails, backs up, and re-reads premise 1 instead (F33).

In F35 she tries to work on the white block, but fails and backs up (F37). She tries again, and succeeds, achieving the result "white red black green" (F40). It must have happened through the translation of premise five, followed by an integration of premise one. In F42 she discovers a contradiction between her partial result and premise 3. This leads to a backup and revision of her mental model to "white green black red" instead (F44). She then integrates premise three into this model, because in F46 she says that the white block is bottommost, a conclusion which only follows from the fact that the green is bottommost but one, combined with the fact that the white is below the green.

The subject then reminds herself that the blue block is still missing from the model and reads premise four which says that the blue block is immediately below the yellow one (F48-F49). There is no evidence that she does anything with this premise. (Since neither the blue nor the yellow block are as yet placed in the model, no extension of the model is possible at this point.) Instead, she reads premise two (F50), and integrates it (F51-F52). After that, the yellow block is part of the model, and premise four can be integrated (F53-F54). Finally, having placed all the objects in the model she reads the question (F55) and derives the answer (F56).

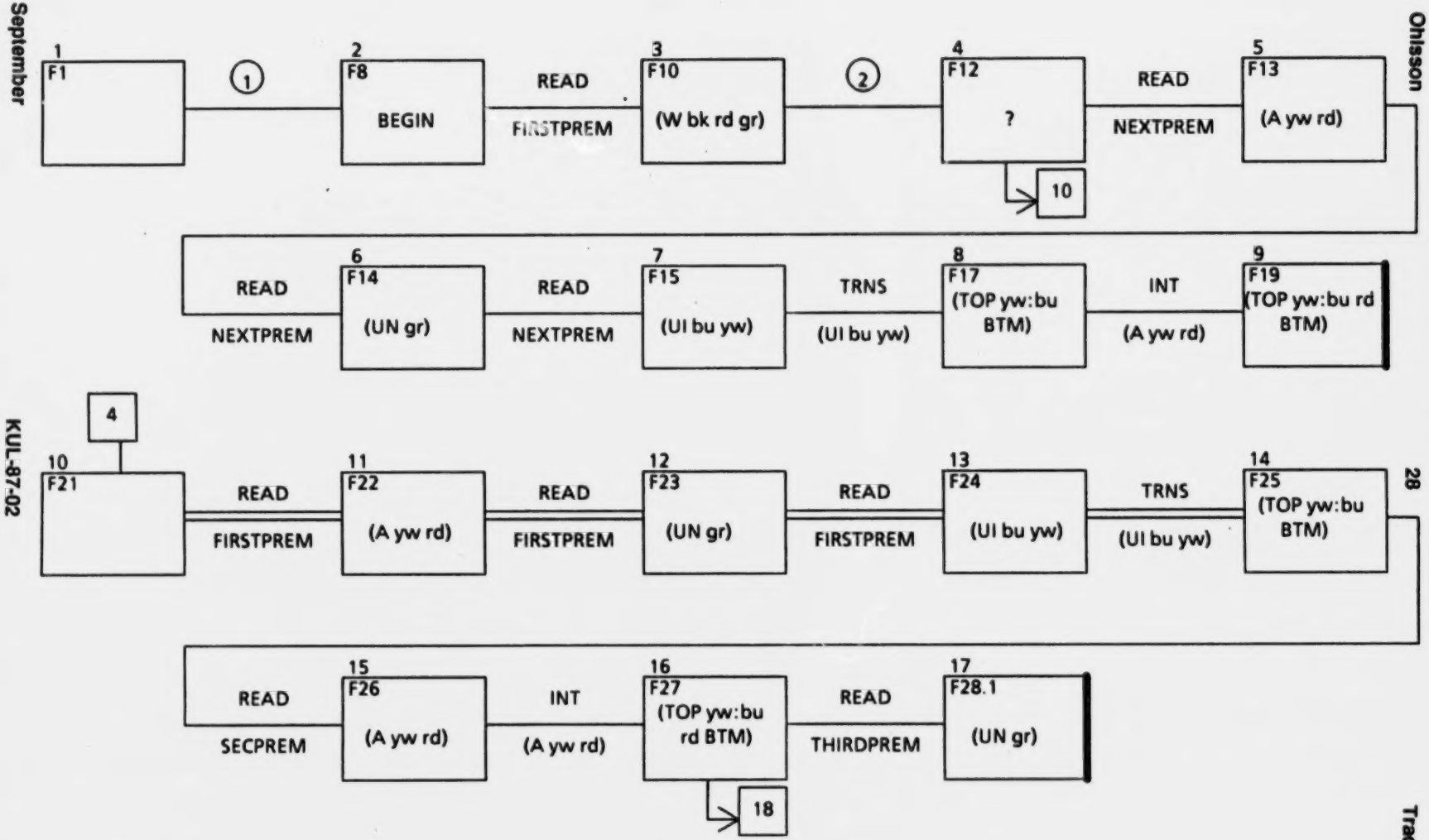

Figure 3-6. Problem Behavior Graph for Subject SI6's Solution Path for the Block Problem, Part1.

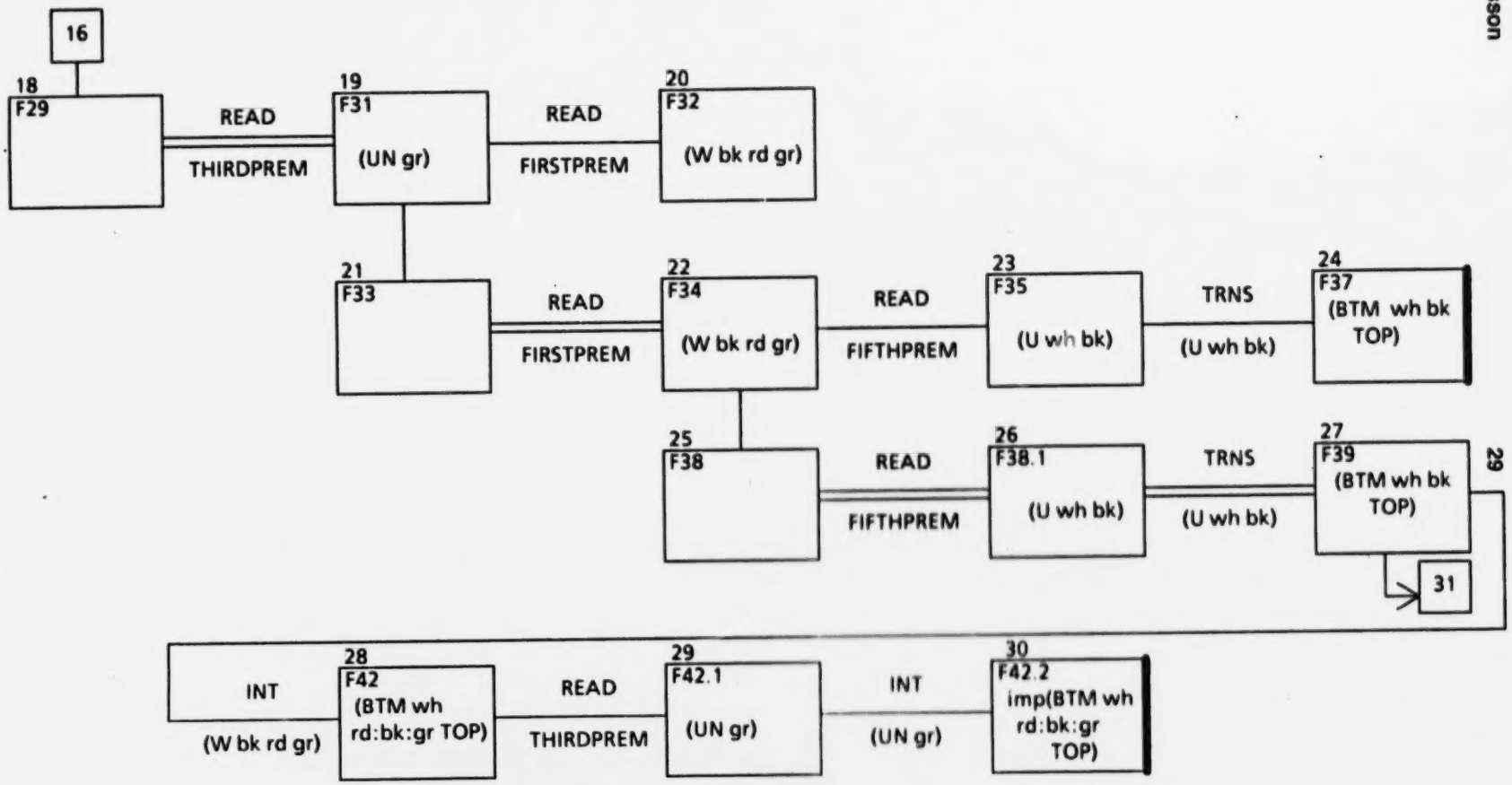

Figure 3-6. Problem Behavior Graph for Subject SI6's Solution Path for the Block Problem, Part2.

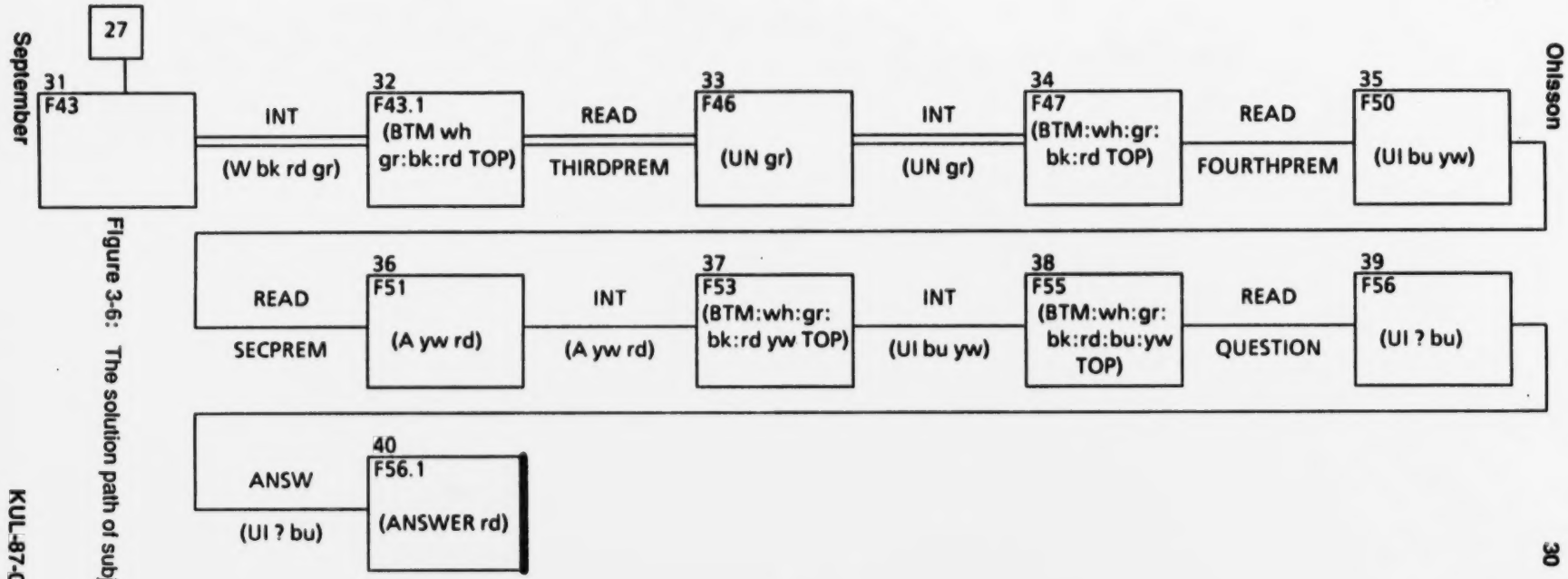

Figure 3-6: The solution path of subject SI6 on the Block Problem.

Figure 3-6. Problem Behavior Graph for Subject SI6's Solution Path for the Block Problem, Part3.
The notation used in this figure is introduced in Figures 3-4 and 3-5.
In order to compress the figure, the following abbreviations are used for the predicate terms:
A = Above, AI = Immediately Above, AM = Topmost, AT = Topmost-But-One
U = Under, UI = Immediately-Under, UM = Bottommost, UN = Bottommost-But-One
I = Adjacent, and W = Between.
The following abbreviations are used for the color terms:
bl = block, bu = blue, gr = green, rd = red, yw = yellow, and wh = white.

The above path hypothesis is summarized graphically in the Problem Behavior Graph (PBG) in Figure 3-6.[19] Since the PBG contains 40 nodes and the solution time was 220 seconds, the *residence time*, i. e., the time the subject spent in each knowledge state before deciding which operator to apply, was 5.5 seconds, a result that is compatible with other analyses of think-aloud protocols (Newell & Simon, 1972).

Special events

Given the above interpretation of the subject's performance, we might ask if the solution path exhibits any remarkable features. Are there any events that are in particular need of explanation, as it were? There are five such events, or groups of events.

First, as the attentive reader has noticed, there is no trace of the partial result "yellow blue red" (which is achieved in fragments F16-F17) in the latter half of the protocol. SI6 creates the ordering "white red black green" (F40) and then continues to integrate the information about the yellow, blue, and red blocks into this ordering, as if she had no previous knowledge of their relative positions. Somewhere in the interval F29-F33 she forgot the mental model she was building. The problem is to explain why such a memory failure occurred at this point, but nowhere else in her solution attempt.

Second, the discovery of the contradiction between her mental model and premise 3 in F42 is crucial for the subject's solution. How did it come about? Premise three happens to be the only premise in the problem which could have shown her that the result achieved in F40 was wrong. What made her re-read this premise at such an appropriate time? Was it a chance event, or was she looking for such information? If she was looking for it, how did she know she needed it?

Third, in the beginning of the solution attempt, the subject *rehearses* premise 2 (F15); in the next pass over the premises, she *re-reads* premise 2 in the corresponding position (F25). In both cases, the assumption of a three-chunk working memory predicts that premise 2 should be available in working memory at that point. Thus, both the rehearsal and the re-reading are in need of explanation.

Fourth, in deriving her first partial result, "yellow blue red", the subject worked with the model from the top and downwards (F16-F17). But later in the protocol, while constructing the sequence "white green black red", she verbalizes her model from the bottom and upwards instead (F40).

---

[19]The notation used in Figure 3-6 is introduced in Figures 3-4 and 3-5. In order to compress the figure, certain abbreviations are used. They are explained in the caption for the third part of the figure.

Fifth, there are four backups in the protocol which are not followed by the exploration of new paths in the problem space, but by repetitions of previously performed inferences (F19, F28.1, F32, and F37). I call them "consolidation backups". There are two questions to be asked about each such event: "Why does it occur when it does?" and "What determines its scope?" (i. e., how many previous steps are repeated?).

3.4. Diagnosing the Subject's Strategy

The solution path (the PBG) is a low-level theory or explanation for the observed performance (the protocol). The next step in the diagnosis is to invent a higher-level theory that explains the solution path. Such an explanation takes the form of a strategy for solving spatial arrangement problems which will generate the hypothesized path when applied to the Blocks Problem.

It would be desirable to mechanize the process of inferring heuristics that explain a particular solution path. Several Artificial Intelligence systems have been proposed that invent a strategy hypothesis, given a protocol (Waterman & Newell, 1971), a problem space (Ohlsson & Langley, 1984, in press), or a solution path (Langley, Ohlsson, & Sage, 1984). Langley, Wogulis, and Ohlsson (this volume) report some recent research with respect to this problem.[20] However, such systems are not yet in practical use, so the practitioner of trace analysis has to be prepared to guess the subject's strategy, and then evaluating his guess by applying it to the path (see next section).

This section hypothesizes a strategy for SI6 and the next section evaluates that hypothesis. I first point out some global properties of SI6's style of problem solving, and then describe her problem solving heuristics in detail.

Global comments on SI6's strategy

There is a strong recency effect in SI6's protocol. The subject's inferences always deal with newly created information. Previous results never seem to confuse her, nor does she make use of them.

There is evidence that the inferential operators TRNS and INT are applied only when certain patterns of information are present. For instance, the subject reads four premises in the beginning of her problem solving attempt before she applies the TRNS operator, apparently waiting for some particular condition to

[20]The reader is referred to Burton (1982) and to Lewis (1986) for examples of systems for the automatic generation of strategy hypotheses that are *not* based on the Enaction Theory.

FPM(<model>)	*Find a proposition related to the current model.* FPM searches working memory for a proposition with one of its arguments already placed in the model. It returns that proposition, if any, or else it fails.
FPP(<proposition>)	*Find a proposition related to a given proposition.* This operator searches working memory for a proposition related in a particular way to the proposition given as argument. It returns that proposition, if any, or else fails. FPP is looking for a *chaining pattern*, i. e., a pair of binary relations such that the second argument of the first proposition is the same as the first argument of the second, e. g., (R x y)(P y z).
GMO(<model>)	*Generate missing objects.* This operator compares the current model and the text, and returns a list of objects which are not yet included in the model. If it cannot find any missing object, it fails.
SCAN(<probe>)	*Scan the text for the element described by the probe.* This operator takes a probe as input, and looks through the text for items that conform to the description in that probe. The probe can be an object, in which case SCAN finds the first premise that mentions that object. The probe can also be the constant UNUSED, in which case SCAN finds the first premise which has not yet participated in any inference. It returns a description of (the location of) the item it finds.
BKUP()	*Backup.* This operator retrieves the knowledge state that was current immediately before the last TRNS or INT inference, and reinstates it as the current knowledge-state.
PREB()	*Prepare for backup.* This operator stores the current knowledge state in long-term memory. It applies immediately before a TRNS or INT inference.

Figure 3-7: Auxiliary problem solving operators for subject SI6.

be satisfied before she starts building her mental model. Similarly, the INT operator is not always applied as soon as there is a new proposition in working memory, but only under certain circumstances. Detailed hypotheses about the patterns she is looking for are stated below.

The subject accesses the external display according to different heuristics during different phases of her problem solving. In the beginning, she is reading the premises in the order in which they are written. After the first application of the TRNS operator she looks around for information which has not yet been used. Finally, at the end, she is searching for information about particular objects.

The subject waits until the end to read the question. This confirms the data-driven character of her processing; a goal-driven system would *begin* with the question.

Formal description of SI6' strategy

In order to describe SI6's strategy as an information processing system, four new operators, two attentional (FPP and FPM) and two perceptual (GMO and SCAN), are needed. They do not change the knowledge state as defined in Figures 3-4 and 3-5, but they control attention, find arguments for the other operators, and access the external display. They are defined in Figure 3-7, which also defines the two backup operators (BKUP and PREB).

The subject's strategy is here represented as a collection of heuristic rules. The rules are stated in a particular format known as a *production system*. In this format each rule has a *condition*, a conjunction of descriptive clauses, and an *action*, a list of problem solving operators. The interpretation of the rule is that if a knowledge state satisfies the condition, then the operators described in the action should be carried out in that state. Production system models are common in the study of human cognition. The reader is referred to Davis and King (1976), Hunt and Poltrock (1974), Klahr, Langley, and Neches (1987b), and Waterman and Hayes-Roth (1978) for general overviews and discussions of production system languages. Although the production system formalism was introduced into psychology in connection with trace analysis (Newell, 1966; Newell & Simon, 1972), there is no inherent conceptual connection between trace analysis and production systems. Other formalisms for the representation of problem solving strategies could be used to express the result of trace analysis.

The production system model of SI6 on the Block Problem is shown in Figure 3-8. The notation used

```
A     <question> <model>    ===> ANSW(<question>)

C1    new(FAIL GMO)  ===>   READ(QUESTION)

I1a   <model> new<proposition> ===> FPM(<model>)(=> proposition);
                                    INT(proposition)

I2    new<model> <proposition>  ===> INT(<proposition>)

T1a   abs<model>  new<proposition>.1 <proposition>.2  ===>
                          FPP(<proposition>.1)(=> proposition);
                          TRNS(<proposition>.1)

B     imp<model>  ===> BKUP()

R3a   new<expression>  (REMAINS = NONE)  ===> GMO(<model>)(=> object);
                                              SCAN(object)(=> premise);
                                              READ(premise)

R2a   new<expression> (HASMODEL = YES)   ===>  SCAN(UNUSED)(=> premise);
                                               READ(premise)

R1    new<expression>  ===>  READ(NEXTPREM)

S1    BEGIN  ===>  READ(FIRSTPREM)
```

Figure 3-8: Production system model of subject SI6.

is a variant of the standard BNF notation.[21] This notation is useful for discussing production systems, because it imposes some discipline on the statement of the production rules while at the same time allowing us to abstract from many of the technical details needed to make a running program. Below I give a natural language paraphrase of, and sometimes a comment to, each production rule.[22]

A *When the question has just been read, and a model is available, try to infer the answer.* The condition on this rule is very general, but SI6 does not read or attend to the question until she is already convinced that the model is complete. Hence, the fact that the question has been attended to is itself an indication that the model is completed, and that the ANSW operator should be applied.

C1 *When there are no more missing objects, read the question.* The fact that there are no more missing objects is a sign that the model is complete and that the problem solving process can move into the question-answering stage.

[21]The rules for this notation can be found in Newell & Simon (1972, pp. 44-46).

[22]The reader need not worry about the somewhat elaborate labeling of the production rules. The labels are intended to facilitate comparison between this production system and other production systems for the same domain in other publications.

I1a *When both a new proposition and a model are available, check if the proposition has the right relation to the model, and, if so, try to integrate it.* The "right relation" is defined by the FPM operator: It returns a proposition that has at least one of its arguments placed in the model.

I2 *When a new model has been derived and there is at least one unused proposition in working memory, then try to integrate that proposition.*

T1a *When there is no model in working memory, but at least one new and one old proposition, then check whether they have the right relation to each other, and, if so, try to translate the most recent of them into a model.* The "right relation" is in this case defined by the FPP operator: It is a chaining pattern like "(x R y)(y Q z)". The three rules I1a, I2, and T1a regulate the effort to draw new inferences from newly created information.

B *When the model contradicts the given information, then back up.*

R3a *When all premises have been used at least once and there is nothing else to do, then find one or more missing objects, locate the premises which deal with those objects and read those premises (regardless of whether they have been read before or not).*

R2a *When a new model has just been achieved and there are still unused premises, then read those premises.*

R1 *When a new model has just been achieved, read the next premise.* The productions R3a, R2a, and R1 represent three different heuristics for how to access the external display.

S1 *Start the problem solving process by reading the first premise.*

In summary, the subject begins by reading the premises in the order in which they are stated in the problem text. When a chaining pattern appears, she starts building a mental model. Having begun building a mental model, she scans the problem text for unused information. Whenever she extends her mental model, she tries to integrate any unused propositional information which is available in working memory. Having considered all premises without completing her model, she identifies specific objects which are missing from the model, and reads any information--old or new--that is available about them. When the model is complete, she reads the question, and answers it by reading off the answer from the model.

3.5. Evaluating the Strategy Hypothesis

The solution path in Figure 3-6 is an hypothesis about the sequence of thoughts the subject had as she solved the Block Problem. The list of production rules in Figure 3-8 is an hypothesis about her problem solving strategy. We do not yet know whether the strategy explains the path or not. A strategy hypothesis must be evaluated by *applying* it to the relevant path. Its justification lies in its ability to generate or reproduce the solution path.

The basic method of applying a production system to a solution path is to ask for each state-operator pair along the path whether there is some production rule which has its condition satisfied in that state and which has that operator as its action. If there is such a rule, that step is *covered* by the production system. If not, the system has made what is known as an *error of omission.* The method is complicated by the fact that several different rules might have their conditions satisfied in one and the same state, and by the fact that the path hypothesis is necessarily incomplete, i. e., it cannot contain all the mental steps the subject actually went through. An explanatory procedure which takes these aspects into account is needed.

In the present analysis, the following procedure was used while applying the production rules in Figure 3-8 to the solution path shown in Figure 3-6. The reader might want to compare this procedure with the discussion in Newell and Simon (1972, pp. 197-199).

1. Suppose that the analysis has proceeded to the *n*th node in the PBG. The step to be explained next is the occurrence of the operator Q leading out from that node. A list is made of all the productions which have such conditions that they could be evoked at that node. The production at the top of the list is assumed to have been evoked. Its action part is compared to the link in the PBG; if it can generate the operator Q, the step leading out from the *n*th node has been explained. The resulting change in the knowledge-state is computed, and the analysis proceeds from the next node.

2. If the action-part of the topmost production cannot generate the operator Q, the protocol is scanned for evidence which contradicts the assumption that the production was fired. If there is no such evidence, the production is assumed to have fired. An node is then interpolated between node *n* and node *(n + 1)*.

3. The process now continues, until either of the following two events occur:
   - The production system finally generates an occurrence of the operator Q, without having contradicted any evidence in the protocol. If this happens, the whole sequence of production occurrences and the corresponding nodes are accepted as part of the solution path. The node which in the PBG appears as the *n*th node, will be

replaced by a sequence of nodes. The first node in the sequence will be identical to the *n*th node, and the last link in the sequence will be the occurrence of the operator Q. The occurrence of the operator has then been explained, and the next node is computed, and the analysis proceeds from it.

- The production system may finally generate some production occurrence which cannot be reconciled with the protocol. Then the entire sequence of production occurrences interpolated after the *n*th node is discarded. The analysis is then resumed at the *n*th node. The topmost production is erased from the list of productions which could have fired at that node. The top-most among those remaining is then assumed to have fired, and the entire process is repeated.

4. If it happens that none of the productions which could have fired at the *n*th node is capable of giving rise to an explanation of the occurrence of the operator Q, the conclusion is that the production system cannot explain what happened at that node. A question mark is entered, the change caused by the operation Q is computed, and the analysis resumes from the *(n+1)*th node.

In order to evaluate how well the production system explains the solution path we have to consider a number of different dimensions, the most important of which are *coverage*, *simplicity*, and *realism*.

Coverage. How many of the knowledge states in the complete solution path are covered by the production rules? There are 48 states, three of which lie outside the problem space. Of the remaining 45 nodes, 42 (93%) are covered. The corresponding figures for the Problem Behavior Graph are 37 and 31 (84%). (The figures differ because the procedure for applying the production system allows the interpolation of states between the nodes in the PBG.)

Another aspect of coverage is the number of special events in the solution path which the account explains. The production system explains the working memory failure in fragment F29-F33. It also explains the discovery of the contradiction in F42. However, the production system does not explain the rehearsal of premise 2 in F15, the re-reading of premise 2 in F25, the change in the order in which the mental model is verbalized, or the occurrence of the two consolidation backups in F28.1 and F32, nor does it explain the scope of any of the consolidation backups.

Simplicity. Taken by itself, an analysis of coverage is not decisive. The problem of coverage can always be solved trivially by adding production rules until every step along the solution path is covered by some rule. In the limit, one could add a separate production rule for each step. Therefore, the drive towards completeness must be balanced by a concern for simplicity.

The number of different productions in Figure 3-8 is 10. The average number of occurrences per production in the complete solution path is 4.8. There are three productions which are used only once: S1, A, and B. S1 and A begin and end a solution process; they fire of necessity only once each. B is the production which causes a backup upon the discovery of a contradiction; it fires only once because the subject discovered a contradiction only once. In short, each production rule adds general explanatory power to the strategy hypothesis, rather than just *ad hoc* coverage of some particular step.

Realism. The production system formalism is a general format for the representation of procedures, but all production rules are not equal, psychologically speaking. In order to be psychologically plausible rules must correspond to pieces of knowledge. The strength of a trace analysis is a function of to what extent it generates weird, complicated, or incomprehensible rules which have no other function than to reproduce the particular observed behavior, and to what extent it generates rules which correspond to useful pieces of heuristic knowledge.

The subjective way of deciding this is to inspect the production system and reflect on each rule, intuiting whether the rule makes sense and whether it is arbitrary. A more intersubjectively valid method is to translate the set of production rules into a running computer program, and then *run the program on other tasks than the one the subject solved.* If the program can solve other tasks, then the production rules are not arbitrary constructions specific to the observed path, but constitute a problem solving strategy of some generality.

The production system in Figure 3-8 was translated into a computer program. The language used was PSS, a production system language designed by the author (Ohlsson, 1979). It shares a family resemblence to such languages as PSG (Newell, 1973), OPS5 (Forgy, 1981), PRISM (Langley, 1983), and ACT (Anderson, 1983). The entire program is reproduced in Appendix A. The program solved the Blocks Problem correctly, generating a solution path which corresponds closely to the solution path by SI6, except for the lack of consolidation backups. In particular, the forgetting of the partial result "yellow blue red" is reproduced by the program, as well as the discovery of the contradiction with the given information in F42. The program was also run on fourteen other spatial arrangement problems of varying difficulty (Ohlsson, 1980a). It solved seven of them correctly. The computer runs showed that the program succeeds on some spatial arrangement problems of equal complexity as the Block Problem, but fails on others. The program also solved 5 out of 6 spatial arrangement problems of lesser complexity, but failed to solve any problems of higher complexity. The main weakness of the program is that it lacks

heuristics for how to proceed when either the FPP or the FPM operator fails. This accounts for the failure on the simpler problems. For the more complex problems, the main source of failure was insufficient working memory capacity. The pattern of results is similar to what one would expect from a human subject.

In summary, the strategy hypothesis does rather well on each of the three basic evaluation dimensions. With respect to coverage, it handles almost all events in the think aloud protocol. The events which are not explained - the rehearsal of premise 2 in F15, the re-reading of premise 2 in F25, the change in how the model is read out, the occurrence and scope of consolidation backups - are all related to working memory capacity. The first-approximation theory of working memory used in this analysis--a box with space for three chunks of information--is, not surprisingly, too course to capture the details of how working memory influenced the problem solving effort. With respect to simplicity, the strategy hypothesis contains no more than ten rules, each of which covers, on the average, five nodes in the path. With respect to realism, computer implementation proved that the strategy can solve other spatial arrangement problems than the one it was designed to solve.

3.6. A Do-It-Yourself Summary

The result of the trace analysis is a description of subject SI6 in terms of her problem space and her problem solving strategy, and a description of her performance in terms of a solution path. The description claims that she successively integrates the propositional information given in the problem text into a mental model of the linear ordering, until the positions of all objects have been determined. Her main difficulty in dealing with the task is that at each point of the process she has to search the problem text for some premise which will enable her to infer the next extension of her model. While carrying out the search through the problem text, the mental model she has achieved up to that point is subject to working memory decay. The major determinant of the shape of her solution effort is not her spatial knowledge, but her strategy for attention allocation.

In order to attempt this kind of cognitive diagnosis the reader should collect a think-aloud protocol from a task he is interested in, and then apply the following explanatory procedures:

1. Identify the subject's problem space:
    a. Construct a representational language for the task by noticing the concepts and representational formats the subject is using in talking about the task.
    b. Define a set of operators based on passages in the protocol which lead to new

results or conclusions.

c. Hypothesize the goal of the subject.

d. Hypothesize a limit on the subject's working memory capacity.

2. Generate a solution path by mapping each fragment in the protocol onto some expression in the representational language. If the expression represents new knowledge about the task, then infer the application of an operator. The solution path is a description of the observed performance *in terms of* the problem space.

3. Invent problem solving heuristics which capture the regularities in the solution path.

4. Evaluate the strategy hypothesis by investigating its coverage, simplicity, and realism.

5. Implement the strategy as a computer program and observe its performance on the experimental task, and on other tasks as well.

4. Implications for Standardized Testing

The process of generating an information processing model with the help of trace analysis is a protracted process involving many decisions and much trial and error on the part of the analyst.[23] Standardized testing, on the other hand, requires that a description of cognitive functioning can be achieved with little enough effort and in short enough a time to be useful in practical contexts. The purpose of this section is to discuss the nature of diagnostic tests that build on information processing concepts, and the role of trace analysis in the construction of such tests.

The psychometric approach to standardized testing is based on the two ideas of measurement and standardization. I analyze these cornerstones of the testing movement in the first two subsections below. The results of trace analysis are, I believe, incompatible with the idea of measurement, but quite compatible with the idea of standardization. I then propose a methodology for the construction of standardized tests based on information processing concepts. This admittedly speculative proposal is called *theory referenced* test construction.

There are, of course, many different bridges to build between the psychometric and the information processing traditions. The reader might want to compare the bridge build here with those constructed by, for example, Carroll (1976), Cooper (1982), Glaser (1986), Hunt (1986), Just and Carpenter (1985), and

[23]The analysis presented in this chapter took approximately six weeks to carry out. The protocol was selected from a corpus of fifty protocols. The analysis of the entire corpus took more than two years.

Snow (1980). A comparative analysis of different conceptualizations of the relation between psychometric and information processing methods would be interesting, but falls outside the scope of the present chapter.

4.1. Trace Analysis and Measurement

The psychometric tradition attempts to describe cognitive functioning with a measure, or, more accurately, a set of measures, defining a point in a multidimensional space (Nunnally, 1967; Sternberg, 1985). But analyses such as the one presented above invalidate this type of description. A set of measures cannot accurately represent the nature of SI6's cognitive processes, for two reasons.

First, the operation of a cognitive mechanism depends essentially on its *structure*. By "structure" I mean the breakdown of the mechanism into parts and the interactions between those parts. For instance, the spatial reasoning of SI6 depends critically on the interaction between her attention allocation and her spatial inferences, as well as on the interaction between her problem solving strategy and her short-term memory capacity. The abstraction involved in expressing her spatial reasoning ability as a measure would inevitably hide those interactions.

Second, the operation of a cognitive mechanism depends essentially on the *content* of its knowledge. The crucial feature of spatial reasoning is not *how many* inference rules a person knows, but exactly *which* rules he knows. The runs with the computer model of SI6 proved that a rule that is necessary for the solution of a one problem may or may not be necessary for the the solution of some other problem at the same level of difficulty (as measured, say, by the number of inferences required to reach the solution). Measures of spatial reasoning ability inevitably abstract from the content of spatial knowledge.

In summary, cognitive mechanisms are not well described by measures. The major implication of information processing concepts with respect to testing is that tests should produce diagnostic descriptions that capture the structure and content of cognitive mechanisms. The complexity of the analysis of subject SI6 raises the question whether this implication is consistent with the notion of standardization.

4.2. Trace Analysis and Standardization

The term "standardized" can be applied either to the behavioral record, to the output description, or to the explanatory procedures of a diagnostic method. It has a different meaning in each case.

The first meaning of standardization is that a test is *a fixed set of problems.* A test consists of problems with known properties that are used over and over again. The practitioner does not need to invent diagnostic problems, he can use existing ones. This is one way in which standardization contributes to practical usefulness. From the point of view of Enaction Theory, generating behavioral records with the help of a fixed set of problems is a great advantage, because the work of constructing a psychologically plausible problem space does not have to be done all over again for each new diagnosis.

The second meaning of standardization is that the purpose of diagnostic inquiry is *to select among pre-defined explanatory accounts.* More accurately, particular diagnoses are instances of well-known explanation patterns. For instance, the names of diseases refer to previously specified physiological states. A doctor who decides that a patient has, say, pneumonia is not discovering a new disease, or inventing a new theory of human physiology, or even constructing a novel account of a patient. He is deciding that his current case is an instance of a known explanation schema. Similarly, a car mechanic who concludes that a car fails to start because of a broken wire is not constructing a theory, but applying a standard explanation type.[24]

Research is our response to a phenomenon that we do not understand. It involves an element of discovery and creative thought precisely because the type of explanation that can account for the phenomenon is not known beforehand, but has to be invented as the explanatory effort proceeds. In a well-understood field of inquiry, on the other hand, we already know which types of explanation will suffice to account for particular types of phenomena. Faced with an instance of a well-understood phenomenon, the task of the practitioner is to select which variant of the relevant explanation type to apply. This is, of course, a much simpler problem than inventing a new explanation type. For example, a medical doctor can diagnose many an infectious disease in a matter of minutes or at most hours, although the research that revealed the physiological mechanism of the disease might have taken many years. In short, the second meaning of "standardized" is that diagnosis does not aim to invent a new explanation, but to select among already known explanations. Diagnostic methods are, by definition, closed methods.

---

[24]Clancey (1985) has developed the difference between *solution construction* and *solution selection* in an A. I. context.

The implication of the above argument is that *standardized testing is only possible in a well-understood domain.* We cannot construct a standardized test for a psychological domain unless we have a theory for human performance in that domain, because the task of a diagnostic procedure is to select among the explanations provided by such a theory. Theory construction must precede test construction, a conclusion already reached by Frederiksen (1986) on the basis of other considerations.[25] This conclusion specifies the role of open methods like trace analysis in test construction: Open methods are needed for the construction of the relevant theory.

The third meaning of standardization is that there exists a *well-specified procedure* for mapping the the set of test responses onto a diagnostic description. One of the great strengths of the psychometric approach is its repertory of well-specified procedures. Statistical theory provides the psychometrician with well motivated, intersubjectively valid algorithms. But the explanatory procedures used in the psychometric approach are based on the idea of measurement, and so cannot be carried over into non-quantitative testing.

In the non-quantitative case diagnosis is a kind of classification (Clancey, 1985). The explanatory procedure classifies the pattern of observed responses as belonging to a particular explanation, or, equivalently, it discriminates between alternative explanations on the basis of the pattern of responses. Recent research in expert systems has shown that complex diagnostic procedures in a variety of domains, including medicine and electronic trouble shooting, can be specified with enough precision to be implemented on a computer (Clancey, 1985; Hayes-Roth, Waterman, & Lenat, 1983). There is, then, reason to believe that procedures for cognitive diagnosis based on information processing concepts can be standardized in the form of computer programs, although there exists to date only a handful of examples (Burton, 1982; Lewis, 1986; Ohlsson & Langley, in press; Sleeman, 1984; Waterman & Newell, 1971).

In summary, the concept of standardization implies (a) that cognitive diagnosis is based on a fixed set of problems, (b) that the purpose of cognitive diagnosis is to select an explanation from a pre-defined set, and (c) that the selection of the explanation is based on a well-specified algorithm. The theories and methods of information processing psychology are quite compatible with these requirements. It should

[25]This conclusion contradicts the idea of using tests as *research instruments*, i. e., as instruments for data collection (rather than for diagnosis). If a theory is a pre-requisite for test construction, then the data required to build that theory must have been collected before the relevant test existed.

therefore be possible to design a methodology for the construction of standardized psychological tests that build on information processing, rather than psychometic, descriptions of mental states.

4.3. Towards Theory Referenced Test Construction

The purpose of this subsection is to outline an admittedly speculative proposal for a methodology that I call *theory referenced test construction*. According to this methodology the construction of a standardized psychological test proceeds through three phases: theory construction, item production, and algorithm design. Each phase will be described in turn.

Theory construction. The construction of a standardized test for diagnosing, say, spatial reasoning, should begin, I propose, with a descriptive investigation of spatial reasoning, using trace analysis and other open and intensive methods that aim for singleton descriptions. The question to be answered by the investigation is "What information processing components (representations, operators, heuristics, goals, inference rules, etc.) have to be postulated to explain a wide variety of human behavior in the relevant task domain?". The results of the investigation are summarized in an information processing theory of human performance in that task domain. The function of that theory is to provide explanations of particular performances. Diagnosis is the process of mapping a particular performance onto the best-fitting explanation.

We can think of a theory of human performance as a space of information processing models. Each model is a specification of an information processing system that can generate (not necessarily correct or efficient) behavior with respect to the relevant task. Each model, i. e., each point in the space, represents a standard (type of) explanation for behavior in the relevant task domain. To explain a particular problem solving performance is to select that model in the space which most closely simulates that performance.

A model space for spatial arrangement problems has been constructed by Ohlsson (1980b, 1982), using trace analysis. A part of this space has been encoded in a *strategy grammar*, a formal device resembling a generative grammar (Ohlsson, 1980a). The model space is defined by a list of information processing components and the rules for how to combine them into particular models. At the most global level of analysis there are several basic *approaches* to spatial arrangement problems. The two most important approaches are the *method of series formation*, which consists in constructing a complete mental model of the linear ordering, and the *method of elimination*, which consists in eliminating all possible answers but one. At the next level of analysis each approach is implemented in several different

problem spaces. For instance, problem spaces for the method of series formation differ with respect to whether the mental model discriminates between adjacent and non-adjacent relations or not, with respect to whether there is an operator for posing hypotheses or not, and so on. (Subject SI6 uses the series formation method, and her problem space--defined in Figures 3-4 and 3-5--contains a symbolic device for discriminating between adjacent and non-adjacent relations, but it does not contain an operator for posing hypotheses.) Each problem space, in turn, can be searched with the help of different *strategies*, each strategy being represented by a set of *heuristics*. For instance, a strategy may or may not include the chaining heuristic (see rule T1a in Figure 3-8). The approaches, problem spaces, and heuristics make up a modeling kit, as it were, out of which particular information processing models can be assembled. To assemble a particular model, one selects an approach, then a problem space which implements that approach, and then a set of heuristics for searching that space. Ohlsson (1982) showed how different subjects can be modeled by different combinations of parts from this space.

The technique of representating a space of information processing models by a modeling kit was first used by Young (1976, 1978) in a study of length seriation in children. He presented a kit of production rules for seriation in which individuals at different levels of development are modeled by a different selection of rules. The same format was used by Young and O'Shea (1981) to describe a model space for multi-column subtraction. Brown and Burton (1978) used a different but related approach to defining a space of models for subtraction. They encoded their space of subtraction models in a structure called a *procedure net*, a network of procedures with calling relations between them. A number of alternative versions of the correct procedure are stored at each node in the procedure net. For instance, there might be several incorrect versions of the borrowing procedure. By making a particular selection among the versions stored at each node in the network, a particular information processing model is assembled, representing a standard explanation for incorrect subtraction answers (a so-called bug). Sleeman (1984) has produced a procedure space for algebra, based on the notion of selecting a set of rules, possibly including some incorrect rules, from a larger set.

Although examples of procedure spaces exist in the literature, they have not yet become common. *The proposal made here is that a procedure space should become a standard way of reporting the results of descriptive studies of human performance.* In particular, I am proposing that a procedure space is the first step in constructing a standardized psychological test. The individual procedures in the space correspond to particular, pre-defined explanations; the task of a diagnostic procedure is to map an

individual onto one of those explanations on the basis of his performance on the test items.

Item production. Given a space of information processing models, the next task of test construction is to produce test items, problems, that will discriminate between those models in the desired way. A problem discriminates between two information processing models A and B if the performance on that problem predicted by model A differs in some observable way from the performance on that problem predicted by model B. The goal of the item production phase is to find a set of problems that discriminates between all members of some given space of models, or that divides the space into equivalence classes.

Item production can be broken down into two processes, item generation and item selection. Both of these processes can be automated. A problem generator is a computer program that can generate possible test items. The art of programming problem generators is currently being explored in research on intelligent tutoring systems (Sleeman & Brown, 1982; Wenger, 1987). In brief, a problem generator needs an analysis of the relevant problem type into fixed and variable parts, and a list of the possible variations. For example, problems of the form "x + y = ?" can be generated by replacing x and y with two random numbers. A problem generator for spatial arrangement problems would be more complicated to program, because it would have to check that the premises it generates make sense when taken together (i. e., that the problem being generated has a solution). A problem generator for, say, electronic trouble shooting would be more complicated still. But problem generators for most tasks that are of interest to test constructors can be programmed with reasonable effort.

After item generation comes item selection. The fact that information processing models are running computer programs can be exploited in order to automate the selection process as well. By running two or more simulation models on a particular problem, one can verify in an intersubjectively valid way whether that problem discriminates between those models or not. Models that generate identical solution paths for that problem are not discriminated, but models that generate different paths are. For instance, spatial arrangement problems that can be solved by integrating the premises in the order in which they are written in the problem text do not discriminate between different strategies for attention allocation, but other problems do. In short, I am proposing that test items should be validated by relating them to the theory of human performance that constitutes the basis for the test. It is this feature of the methodology proposed here that motivates the term "theory referenced test construction".

Item production can be fully automated by interleaving item generation and item selection. A

computer system for item production would generate an item, run the relevant models on it, and decide whether to keep the item on the basis of whether it discriminates between those models. The cycle of problem generation and model running would continue until the system has found a set of items that makes the desired discriminations between all the relevant models. That set of problems is then a test for whatever aspect of human cognition is described by that space of models.

Algorithm design. The relationship between a pattern of responses on a test, on the one hand, and a space of information processing models, on the other, can be very complex. If a test is to be useful in practical contexts, it must be possible to design an algorithm that quickly selects that model which best accounts for any particular pattern of responses. In principle, a pattern classifier consists of a discrimination tree that makes successive decisions depending upon the answers to each diagnostic item. The highly successful DEBUGGY system for classification of subtraction errors (Burton, 1982), and the construction of expert systems for medical diagnosis, electronic trouble shooting, and similar domains (Clancey, 1985) show that complex pattern classification algorithms can be designed and programmed.

Admittedly, the methodology for test construction outlined here cannot compete with the psychometric approach with respect to the processing of test responses. Given the psychometric idea of describing a mental state as a point in a multi-dimensional space, standard statistical techniques can be used to process the data from any test, regardless of the problems in the test, regardless of what the test measures, and even regardless of changes in the underlying theory, e. g., changes in the assumptions about how many distinct abilities there are. In contrast, the methodology outlined here requires that a new classification algorithm is designed for each new test.

In summary, theory referenced test construction proceeds by (a) constructing a space of information processing models, each model describing a possible state of mind, (b) producing a test, I. e., a set of items that can discriminate between those models, and (c) designing a pattern classification algorithm that selects the best-fitting model for a particular set of responses to the test items.

The above proposal is admittedly speculative. But the two last phases of the proposed methodology--item production and algorithm design--rely on standard programming techniques. No conceptual advances are needed to realize those two phases of the methodology. The speculative nature of the proposal comes to the fore in the first step. It is not obvious that we know how to construct model spaces that simulate people with enough accuracy to be used as bases for test construction. The example

provided by research on subtraction skills is encouraging (Brown & Burton, 1978; Burton, 1982). Furthermore, our ability to construct such model spaces is a function of the quality of our psychological theories. Presumably, continued psychological research will lead to better and more accurate theories of human cognition, and the better our theories, the more feasible the methodology of theory referenced test construction.

Reference Notes

Newell, A. (1987). Unified theories of cognition. *The William James Lecture Series*, Harvard University.

References

Anderson, J. R., (Ed.)(1981). *Cognitive skills and their acquisition.* San Francisco: Freeman.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, J. R., Boyle, D. F., & Reiser, B. J. (1985). Intelligent tutoring systems. *Science, 228*, pp. 456-462.

Bolc, L., (Ed.)(1987). *Computational models of learning.* Berlin: Springer-Verlag.

Brown, J. S., & Burton, R. R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science, 2*, 155-192.

Burton, R. (1982). Diagnosing bugs in a simple procedural skill. In D. Sleeman & J. S. Brown (Eds.), *Intelligent tutoring systems*, (pp. 157-183). London: Academic Press.

Carroll, J. B. (1976). Psychometric tests as cognitive tasks: A new "structure of intellect". In L. B. Resnick (Ed.), *The Nature of intelligence*, (pp. 27-56). Hillsdale, NJ: Erlbaum.

Clancey, W. J. (1985, June). *Heuristic classification.* (Tech. Rep, No. STAN-CS-85-1066). Stanford, CA: Stanford University, Department of Computer Science.

Cooper, L. A. (1982). Strategies for visual comparison and representation: Individual differences. In R. J. Sternberg, (Ed.). *Advances in the psychology of human intelligence. Volume 1*, pp. 77-124. Hillsdale, NJ: Erlbaum.

Davis, R., & King, J. (1976). An overview of production systems. In E. W. Elcock & D. Michie (Eds.), *Machine Intelligence. Vol 1.* New York: Wiley.

Dukes, N. F. (1968). N = 1. *Psychological Bulletin, 64*, 74-79.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis. Verbal reports as data.* Cambridge, MA: The MIT Press.

Forgy, C. L. (1981). *OPS5 User's Manual.* (Tech. Rep. No. CMU-CS-81-135). Pittsburgh, PA: Carnegie-Mellon University, Computer Science Department.

Frederiksen, N. (1984). Implications of cognitive theory for instruction in problem solving. *Review of Educational Research, 54*(3), 363-407.

Frederiksen, N. (1986). Construct validity and construct similarity: Methods for use in test development and test validation. *Multivariate Behavioral Research, 21*, 3-28.

Gardner, H. (1985). *The mind's new science. A history of the cognitive revolution.* New York: Basic Books.

Gentner, D., & Stevens, A. L. (Eds.)(1983). *Mental models.* Hillsdale, NJ: Erlbaum.

Glaser, R. (1986). The integration of instruction and testing. *Redesign of testing for the 21st century. Proceedings from the ETS Invitational Conference*, October 1985. Princeton, NJ: ETS.

Hagert, G. (1980a). *Cognitive processing in a spatial series task: A cognitive simulation study.* Working paper from the Cognitive Seminar (ISSN 0349-5124), University of Stockholm, No. 7.

Hagert, G. (1980b). Stability and change in strategies: Three simulations of one subject with one-year intervals. *Proceedings of the fourth Conference on Artificial Intelligence and Simulation of Behavior*, Amsterdam.

Hagert, G., & Rollenhagen, C. (1981). *Learning in a spatial series task: A production system analysis.* Working paper from the Cognitive Seminar (ISSN 0349-5124), University of Stockholm, No. 14.

Hayes-Roth, F., Waterman, D. A., & Lenat, D. B. (1983). *Building expert systems.* Reading, MA: Addison-Wesley.

Holsti, O. R. (1968). Content analysis. In G. Lindzey, & E. Aronson (Eds.), *The Handbook of Social Psychology, Second ed. Volume Two. Research Methods* (pp. 596-692). Reading, MA: Addison-Wesley.

Hunt, E. B. (1986). Cognitive research and future test design. *Redesign of testing for the 21st century. Proceedings of the ETS Invitational Conference*, (pp. 9-23), October 1985. Princeton, NJ: ETS.

Hunt, E. B., & Poltrock, S. E. (1974). The mechanics of thought. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition.* New York, NJ: Wiley.

Johnson-Laird, P. N. (1983). *Mental models.* Cambridge, MA: Harvard University Press.

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review, 92*, 137-171.

Klahr, D., Langley, P., & Neches, R. (Eds.) (1987a). *Production system models of learning and development.* Cambridge, MA: MIT Press.

Klahr, D., Langley, P., & Neches, R. (1987b). Learning, development, and production systems. In Klahr, D., Langley, P., & Neches, R. (Eds.) *Production system models of learning and development.* Cambridge, MA: MIT Press.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology, 17*, 248-294.

Langley, P. (1983). Exploring the space of cognitive architectures. *Behavior Research Methods & Instrumentation, 15*(2), 289-299.

Langley, P., Ohlsson, S., & Sage, S. (1984). *A machine learning approach to student modeling* (Tech. Rep. No. CMU-RI-TR-84-7). Pittsburgh, PA: Carnegie-Mellon University.

Lewis, C. M. (1986). *Intention-based diagnosis of novice programming errors.* Los Altos, Calif.: Kaufmann.

Maier, N. R. F., (1970). *Problem solving and creativity in individuals and groups.* Belmont, CA: Brooks/Cole.

Newell, A. (1966, June). *On the analysis of human problem solving protocols.* (Tech. Rep.). Pittsburgh, PA: Carnegie Institute of Technology.

Newell, A. (1973). Production systems: Models of control structures. In W. G. Chase (Ed.), *Visual information processing* (pp. 463-526). New York: Academic Press.

Newell, A. (1980) Reasoning, problem solving and decision processes: The problem space as a

fundamental category. In R. Nickerson (Ed.), *Attention and performance VIII.* Hillsdale, NJ: Erlbaum.

Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review, 65*(3), 151-166.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Nunnally, J. (1967). *Psychometric Theory.* New York: McGraw-Hill.

Ohlsson, S. (1979, September). *PSS3 Reference Manual.* (Working Papers from the Cognitive Seminar No. 4). Stockholm, Sweden: University of Stockholm, Department of Psychology.

Ohlsson, S. (1980a, January). *Competence and reasoning with common spatial concepts.* (Working Papers from the Cognitive Seminar No. 6). Stockholm, Sweden: University of Stockholm, Department of Psychology.

Ohlsson, S. (1980b, July). Strategy grammars. An approach to generality in computer simulation of human reasoning. *Proceedings of the AISB-80 Conference on Artificial Intelligence*, Amsterdam.

Ohlsson, S. (1982). *Problem solving strategies in a reasoning task.* (Tech. Rep. No. 340). Uppsala, Sweden: University of Uppsala, Department of Psychology.

Ohlsson, S. (1983). The enaction theory of thinking and its educational implications. *Scandinavian Journal of Educational Research, 27*, 73-88.

Ohlsson, S. (1984a, June). Attentional heuristics in human thinking. *Proceedings of the Sixth Conference of the Cognitive Science Society.* Boulder, CO.

Ohlsson, S. (1984b). Induced strategy shifts in spatial reasoning. *Acta Psychologica, 57*, 47-67.

Ohlsson, S. (1984c). Restructuring revisited II. An information processing theory of restructuring and insight. *Scandinavian Journal of Psychology, 25*, 117-129.

Ohlsson, S. (In press). Computer simulation and its impact on educational research and practice. *International Journal of Educational Research.*

Ohlsson, S., & Langley, P. (1984). Towards automatic discovery of simulation models. Proceedings from the Sixth European Conference of Artificial Intelligence, Pisa, Italy

Ohlsson, S., & Langley, P. (In press). Psychological evaluation of path hypothesis in cognitive diagnosis. In Mandl, H. & Lesgold, A. (Eds.) *Learning issues for intelligent tutoring systems.* New York: Springer

Rosenbloom, P. S., Laird, J. E., McDermott, J., Newell, A., & Orciuch, E. (1985). *R1-Soar: An experiment in knowledge-intensive programming in a problem-solving architecture* (Tech. Rep. No. CMU-CS-85-110). Pittsburgh, PA: Carnegie-Mellon University.

Sleeman, D. (1984). An attempt to understand student's understanding of basic algebra. *Cognitive Science, 8*(4), 387-412.

Sleeman, D., & Brown, J. S., (Eds.), (1982). *Intelligent tutoring systems.* London, U. K.: Academic Press.

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology, 7* pp. 268-288.

Snow, R. E. (1980). Aptitude processes. In: R. E. Snow, P. A. Frederico, & W. E. Montague (Eds.),

Aptitude, learning, and instruction. Volume 1: Cognitive process analyses of aptitude (pp. 27-63). Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1985). General intellectual ability. In R. J. Sternberg (Ed.), *Human abilities. An information processing approach* (pp. 5-30). New York, NY: Freeman.

Toulmin, S. (1972). *Human understanding. Vol. 1. General introduction and Part I.* Oxford, U. K.: Oxford University Press.

Waterman, D. A., & Hayes-Roth, F. (1978). An overview of pattern-directed inference systems. In D. A. Waterman & F. Hayes-Roth (Ed.), *Pattern-directed inference systems.* New York, NY: Academic Press.

Waterman, D. A., & Newell, A. (1971). Protocol analysis as a task for artificial intelligence. *Artificial Intelligence, 2,* 285-318.

Wenger, E. (1987). *Artificial intelligence and tutoring systems.* Los Altos, Calif.: Kaufmann.

Williams, M. D., & Santos-Williams, S. (1980). Method for exploring retrieval processes using verbal protocols. In R. S. Nickerson (Ed.), *Attention and performance. Vol 8.* Hillsdale, NJ: Erlbaum.

Williams, M. D., & Hollan, J. D. (1981). The process of retrieval from very-long term memory. *Cognitive Science, 5,* 87-119.

Young, R. M. (1976). *Seriation in children. An artificial intelligence analysis.* Basel, West Germany: Birkhauser Verlag.

Appendix A. Simulation Program for S16

The following is a runnable simulation model of subject S16. It consists of the production rules in Figure 3-8, written in a computer implemented production system language called PSS (Ohlsson, 1979).

```
(P0 (ANSWER X1) ===> SAY(X1);
                     STOPALL)

(P1 (NEW <QSTN>) <MODEL> ===> UNMK((NEW <QSTN>));
                              GOTO(ANSW))

(P2 (NEW (FAIL GMO)) ===> UNMK((NEW (FAIL GMO)));
                          READ(QUESTION))

(P3A (NTC: <PROP>) <MODEL> ===> GOTO (INT))

(P3B (NEW <PROP>) <MODEL> ===> GOTO(FPM))

(P4 (NEW <MODEL>) <PROP> ===> UNMK((NEW <MODEL>));
                              MARK(<PROP> ; NTC:);
                              GOTO(INT))

(P5A (ABS <MODEL>) (NTC: <PROP>.1) <PROP>.2 ===>
                              UNMK((NTC: <PROP>.1));
                              RHRS(<PROP>.2);
                              RHRS(<PROP>.1);
                              GOTO(TRNS))

(P5B (ABS <MODEL>) (NEW <PROP>.1) <PROP>.2 ===> GOTO(FPP))

(P6 (IMP <MODEL>) ===>  BKUP())

(P7A (NEW <EXPRESSION>) (MISSING: (X1)) ===>
                              UNMK((NEW <EXPRESSION>));
                              DEL((MISSING: (X1)));
                              SCAN((X1)) (=> PREMISE);
                              READ(PREMISE))

(P7B (NEW <EXPRESSION>) (MISSING: (X1 <SEQ>)) ===>
                              UNMK((NEW <EXPRESSION>));
                              REPL((MISSING: X1 <SEQ>));
                                   (MISSING: (<SEQ>)));
                              SCAN((X1)) (=> PREMISE);
                              READ(PREMISE))

(P7C (NEW <EXPRESSION>) (REMAINS = NONE) ===>
                              UNMK((NEW <EXPRESSION>));
                              NTC(<MODEL>);
                              GMO(<MODEL>) (=> LIST);
                              MARK(<EXPRESSION> ; NEW);
                              INS((MISSING: LIST)))
```

Appendix A. Cont'd

```
(P8A  (NEW <EXPRESSION>) (UNUSED: (X1)) ===>
                                 UNMK((NEW <EXPRESSION>));
                                 DEL((UNUSED: (X1)));
                                 READ(X1))

(P8B (NEW <EXPRESSION>) (UNUSED: (X1 <SEQ>)) ===>
                                 UNMK((NEW<EXPRESSION>));
                                 REPL((UNUSED: (X1 <SEQ>)) ;
                                      (UNUSED: (<SEQ.>)));
                                 READ(X1))

(P8C (NEW <EXPRESSION>) <MODEL> ===>
                                 SCAN(UNUSED)(=> LIST);
                                 INS((UNUSED: LIST)))

(P9 (NEW <EXPRESSION>) ===> UNMK((NEW <EXPRESSION>));
                            READ(NEXTPREM))

(P10 BEGIN ===> READ(FIRSTPREM))
```

KUL Reports

1985

Ohlsson, S., & Langley, P. (April, 1985). Psychological evaluation of path hypotheses in cognitive diagnosis (Technical Report No. 1985/2). Pittsburgh: Learning Research and Development Center, University of Pittsburgh.

1986

Ohlsson, S. (January, 1986). Some principles of intelligent tutoring (Technical Report No. 1986/2). Pittsburgh: Learning Research and Development Center, University of Pittsburgh.

Ohlsson, S. (June, 1986). Computer simulation and its impact on educational research and practice (Technical Report No. 1986/14). Pittsburgh: Learning Research and Development Center, University of Pittsburgh.

Ohlsson, S. (October, 1986). Sense and reference in the design of interactive illustrations for rational numbers (Technical Report No. 1986/18). Pittsburgh: Learning Research and Development Center, University of Pittsburgh.

1987

Ohlsson, S. (April, 1987). A semantics for fraction concepts (Technical Report No. KUL-87-01). Pittsburgh: Learning Research and Development Center, University of Pittsburgh.

Ohlsson, S. (September, 1987). Trace analysis and spatial reasoning: An example of intensive cognitive diagnosis and its implications for testing (Technical Report No. KUL-87-02). Pittsburgh: Learning Research and Development Center, University of Pittsburgh.

Ohlsson, S., Nicholas, S., & Bee, N. (December, 1987). Interactive illustrations for fractions: A progress report (Technical Report No. KUL-87-03). Pittsburgh: Learning Research and Development Center, University of Pittsburgh.

Ohlsson, S., & Rees, E. (December, 1987). Rational learning: Deriving procedures from state constraints (Technical Report No. KUL-87-04). Pittsburgh: Learning Research and Development Center, University of Pittsburgh.

1988

Ohlsson, S. (February, 1988). Mathematical meaning and applicational meaning in the semantics for fractions and related concepts (Technical Report No. KUL-88-01). Pittsburgh: Learning Research and Development Center, University of Pittsburgh.

Table of Contents

1. On Methodology 4
2. The Enaction Theory and Trace Analysis 8
   2.1. The Enaction Theory of Thinking 9
   2.2. The Method of Trace Analysis 11
3. Trace Analysis Applied to Spatial Reasoning 14
   3.1. The Subject and the Behavioral Record 17
   3.2. Diagnosing the Subject's Problem Space 19
   3.3. Diagnosing the Subject's Solution Path 25
   3.4. Diagnosing the Subject's Strategy 32
   3.5. Evaluating the Strategy Hypothesis 37
   3.6. A Do-It-Yourself Summary 40
4. Implications for Standardized Testing 41
   4.1. Trace Analysis and Measurement 42
   4.2. Trace Analysis and Standardization 43
   4.3. Towards Theory Referenced Test Construction 45

List of Figures

Figure 3-1: Examples of spatial arrangement problems. 15
Figure 3-2: Think-aloud protocol from SI6 on the Block Problem, Part 1 18
Figure 3-3: Think-aloud protocol from subject SI6 on the Block Problem, Part 2. 19
Figure 3-4: Mental representation of subject SI6 for the Block Problem. 20
Figure 3-5: Basic problem solving operators of subject SI6. 24
Figure 3-6: The solution path of subject SI6 on the Block Problem. 30
Figure 3-7: Auxiliary problem solving operators for subject SI6. 33
Figure 3-8: Production system model of subject SI6. 35

1987/03/09

Distribution List [Pittsburgh/Lesgold] NR 4422539

Dr. Phillip L. Ackerman
University of Minnesota
Department of Psychology
Minneapolis, MN 55455

Dr. Beth Adelson
Department of Computer Science
Tufts University
Medford, MA 02155

Air Force Human Resources Lab
AFHRL/MPD
Brooks AFB, TX 78235

AFOSR,
Life Sciences Directorate
Bolling Air Force Base
Washington, DC 20332

Technical Director,
Army Human Engineering Lab
Aberdeen Proving Ground
MD 21005

Dr. Robert Ahlers
Code N711
Human Factors Laboratory
Naval Training Systems Center
Orlando, FL 32813

Dr. Ed Aiken
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. John Allen
Department of Psychology
George Mason University
4400 University Drive
Fairfax, VA 22030

Dr. James Anderson
Brown University
Center for Neural Science
Providence, RI 02912

Dr. John R. Anderson
Department of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213

Dr. Nancy S. Anderson
Department of Psychology
University of Maryland
College Park, MD 20742

Technical Director, ARI
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Alan Baddeley
Medical Research Council
Applied Psychology Unit
15 Chaucer Road
Cambridge CB2 2EF
ENGLAND

Dr. Patricia Baggett
University of Colorado
Department of Psychology
Box 345
Boulder, CO 80309

Dr. Eva L. Baker
UCLA Center for the Study
of Evaluation
145 Moore Hall
University of California
Los Angeles, CA 90024

Dr. Meryl S. Baker
Navy Personnel R&D Center
San Diego, CA 92152-6800

prof. dott. Bruno G. Bara
Unita di ricerca di
intelligenza artificiale
Universita di Milano
20122 Milano - via F. Sforza 23
ITALY

Dr. William M. Bart
University of Minnesota
Dept. of Educ. Psychology
330 Burton Hall
178 Pillsbury Dr., S.E.
Minneapolis, MN 55455

Dr. Jackson Beatty
Department of Psychology
University of California
Los Angeles, CA 90024

1987/03/09

Distribution List [Pittsburgh/Lesgold] NR 4422539

Dr. John Black
Teachers College
Columbia University
525 West 121st Street
New York, NY 10027

Dr. Jeff Bonar
Learning R&D Center
University of Pittsburgh
Pittsburgh, PA 15260

Dr. Gordon H. Bower
Department of Psychology
Stanford University
Stanford, CA 94306

Dr. Robert Breaux
Code N-095R
Naval Training Systems Center
Orlando, FL 32813

Commanding Officer
CAPT Lorin W. Brown
NROTC Unit
Illinois Institute of Technology
3300 S. Federal Street
Chicago, IL 60616-3793

Dr. John S. Brown
XEROX Palo Alto Research
Center
3333 Coyote Road
Palo Alto, CA 94304

Maj. Hugh Burns
AFHRL/IDE
Lowry AFB, CO 80230-5000

Dr. Jaime Carbonell
Carnegie-Mellon University
Department of Psychology
Pittsburgh, PA 15213

Dr. Pat Carpenter
Carnegie-Mellon University
Department of Psychology
Pittsburgh, PA 15213

LCDR Robert Carter
Office of the Chief
of Naval Operations
OP-01B
Pentagon
Washington, DC 20350-2000

Chair, Department of
Psychology
College of Arts and Sciences
Catholic University of
America
Washington, DC 20064

Dr. Michelene Chi
Learning R & D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

Dr. L. J. Chmura
Computer Science and Systems
Code: 7590
Information Technology Division
Naval Research Laboratory
Washington, DC 20375

Mr. Raymond E. Christal
AFHRL/MOE
Brooks AFB, TX 78235

Assistant Chief of Staff
for Research, Development,
Test, and Evaluation
Naval Education and
Training Command (N-5)
NAS Pensacola, FL 32508

Dr. Allan M. Collins
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02138

Dr. Stanley Collyer
Office of Naval Technology
Code 222
800 N. Quincy Street
Arlington, VA 22217-5000

Brian Dallman
3400 TTW/TTGXS
Lowry AFB, CO 80230-5000

1987/03/09

Distribution List [Pittsburgh/Lesgold] NR 4422539

Dr. Diane Damos
Arizona State University
Department of Psychology
Tempe, AZ 85287

Dr. Denise Dellarosa
Department of Psychology
Yale University
Box 11A, Yale Station
New Haven, CT 06520

Dr. R. K. Dismukes
Associate Director for Life Sciences
AFOSR
Bolling AFB
Washington, DC 20332

Dr. Stephanie Doan
Code 6021
Naval Air Development Center
Warminster, PA 18974-5000

Dr. Emanuel Donchin
University of Illinois
Department of Psychology
Champaign, IL 61820

Defense Technical
Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC
(12 Copies)

Dr. Susan Embretson
University of Kansas
Psychology Department
426 Fraser
Lawrence, KS 66045

Dr. Randy Engle
Department of Psychology
University of South Carolina
Columbia, SC 29208

Dr. William Epstein
University of Wisconsin
W. J. Brogden Psychology Bldg.
1202 W. Johnson Street
Madison, WI 53706

ERIC Facility-Acquisitions
4833 Rugby Avenue
Bethesda, MD 20014

Dr. K. Anders Ericsson
University of Colorado
Department of Psychology
Boulder, CO 80309

Dr. Martha Farah
Department of Psychology
Carnegie-Mellon University
Schenley Park
Pittsburgh, PA 15213

Dr. Beatrice J. Farr
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Marshall J. Farr
Farr-Sight Co.
2520 North Vernon Street
Arlington, VA 22207

Dr. Paul Feltovich
Southern Illinois University
School of Medicine
Medical Education Department
P.O. Box 3926
Springfield, IL 62708

Dr. Craig I. Fields
ARPA
1400 Wilson Blvd.
Arlington, VA 22209

J. D. Fletcher
9931 Corsica Street
Vienna VA 22180

Dr. Kenneth D. Forbus
University of Illinois
Department of Computer Science
1304 West Springfield Avenue
Urbana, IL 61801

Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138

1987/03/09

Distribution List [Pittsburgh/Lesgold] NR 4422539

Dr. Alfred R. Fregly
AFOSR/NL
Bolling AFB, DC 20332

Dr. Michael Friendly
Psychology Department
York University
Toronto ONT
CANADA M3J 1P3

Julie A. Gadsden
Information Technology
Applications Division
Admiralty Research Establishment
Portsdown, Portsmouth PO6 4AA
UNITED KINGDOM

Dr. Michael Genesereth
Stanford University
Computer Science Department
Stanford, CA 94305

Dr. Dedre Gentner
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

Dr. Lee Giles
AFOSR
Bolling AFB
Washington, DC 20332

Dr. Robert Glaser
Learning Research
& Development Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

Dr. Marvin D. Glock
13 Stone Hall
Cornell University
Ithaca, NY 14853

Dr. Sam Glucksberg
Department of Psychology
Princeton University
Princeton, NJ 08540

Dr. Daniel Gopher
Industrial Engineering
& Management
TECHNION
Haifa 32000
ISRAEL

Dr. Sherrie Gott
AFHRL/MODJ
Brooks AFB, TX 78235

Dr. T. Govindaraj
Georgia Institute of Technology
School of Industrial & Systems
Engineering
Atlanta, GA 30332

Dr. Richard H. Granger
Department of Computer Science
University of California, Irvine
Irvine, CA 92717

Dr. James G. Greeno
University of California
Berkeley, CA 94720

Dr. Henry M. Halff
Halff Resources, Inc.
4918 33rd Road, North
Arlington, VA 22207

Dr. Bruce Hamill
The Johns Hopkins University
Applied Physics Laboratory
Laurel, MD 20707

Dr. John M. Hammer
Center for Man-Machine
Systems Research
Georgia Institute of Technology
Atlanta, GA 30332

Dr. Ray Hannapel
Scientific and Engineering
Personnel and Education
National Science Foundation
Washington, DC 20550

Dr. Harold Hawkins
Office of Naval Research
Code 1142CS
800 N. Quincy Street
Arlington, VA 22217-5000

1987/03/09

Distribution List [Pittsburgh/Lesgold] NR 4422539

Dr. Barbara Hayes-Roth
Department of Computer Science
Stanford University
Stanford, CA 95305

Dr. Frederick Hayes-Roth
Teknowledge
525 University Ave.
Palo Alto, CA 94301

Dr. Joan I. Heller
505 Haddon Road
Oakland, CA 94606

Dr. Geoffrey Hinton
Carnegie-Mellon University
Computer Science Department
Pittsburgh, PA 15213

Dr. James D. Hollan
MCC,
Human Interface Program
3500 West Balcones Center Dr.
Austin, TX 78759

Dr. John Holland
University of Michigan
2313 East Engineering
Ann Arbor, MI 48109

Dr. Melissa Holland
Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Robert W. Holt
Department of Psychology
George Mason University
4400 University Drive
Fairfax, VA 22030

Ms. Julia S. Hough
Lawrence Erlbaum Associates
6012 Greene Street
Philadelphia, PA 19144

Dr. James Howard
Dept. of Psychology
Human Performance Laboratory
Catholic University of
America
Washington, DC 20064

Dr. Earl Hunt
Department of Psychology
University of Washington
Seattle, WA 98105

Dr. Ed Hutchins
Intelligent Systems Group
Institute for
Cognitive Science (C-015)
UCSD
La Jolla, CA 92093

Dr. Janet Jackson
Rijksuniversiteit Groningen
Biologisch Centrum, Vleugel D
Kerklaan 30, 9751 NN Haren (Gn.)
NETHERLANDS

Dr. R. J. K. Jacob
Computer Science and Systems
Code: 7590
Information Technology Division
Naval Research Laboratory
Washington, DC 20375

Dr. Zachary Jacobson
Bureau of Management Consulting
365 Laurier Avenue West
Ottawa, Ontario K1A 0S5
CANADA

Pharm.-Chim. en Chef Jean Jacq
Division de Psychologie
Centre de Recherches du
Service de Sante des Armees
108 Boulevard Pinel
69272 Lyon Cedex 03, FRANCE

Dr. Robert Jannarone
Department of Psychology
University of South Carolina
Columbia, SC 29208

Dr. Claude Janvier
Directeur, CIRADE
Universite' du Quebec a Montreal
P.O. Box 8888, St. "A"
Montreal, Quebec H3C 3P8
CANADA

1987/03/09

Distribution List [Pittsburgh/Lesgold] NR 4422539

COL Dennis W. Jarvi
Commander
AFHRL
Brooks AFB, TX 78235-5601

Dr. Robin Jeffries
Hewlett-Packard Laboratories
P.O. Box 10490
Palo Alto, CA 94303-0971

Dr. Douglas H. Jones
Thatcher Jones Associates
P.O. Box 6640
10 Trafalgar Court
Lawrenceville, NJ 08648

Dr. Marcel Just
Carnegie-Mellon University
Department of Psychology
Schenley Park
Pittsburgh, PA 15213

Dr. Daniel Kahneman
Department of Psychology
University of California
Berkeley, CA 94720

Dr. Milton S. Katz
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Steven W. Keele
Department of Psychology
University of Oregon
Eugene, OR 97403

Dr. Wendy Kellogg
IBM T. J. Watson Research Ctr.
P.O. Box 218
Yorktown Heights, NY 10598

Dr. David Kieras
University of Michigan
Technical Communication
College of Engineering
1223 E. Engineering Building
Ann Arbor, MI 48109

Dr. Walter Kintsch
Department of Psychology
University of Colorado
Campus Box 345
Boulder, CO 80302

Dr. David Klahr
Carnegie-Mellon University
Department of Psychology
Schenley Park
Pittsburgh, PA 15213

Mr. Al Kleider
Army Research Office
P.O. Box 12211
Research Triangle Park
North Carolina 27709-2211

Dr. Ronald Knoll
Bell Laboratories
Murray Hill, NJ 07974

Dr. Stephen Kosslyn
Harvard University
1236 William James Hall
33 Kirkland St.
Cambridge, MA 02138

Dr. Kenneth Kotovsky
Department of Psychology
Community College of
Allegheny County
800 Allegheny Avenue
Pittsburgh, PA 15233

Dr. David H. Krantz
2 Washington Square Village
Apt. # 15J
New York, NY 10012

Dr. Patrick Kyllonen
325 Aderhold
Department of Educational
Psychology
University of Georgia
Athens, GA 30602

Dr. David R. Lambert
Naval Ocean Systems Center
Code 441T
271 Catalina Boulevard
San Diego, CA 92152-6800

1987/03/09

Distribution List [Pittsburgh/Lesgold] NR 4422539

Dr. Jill Larkin
Carnegie-Mellon University
Department of Psychology
Pittsburgh, PA 15213

Dr. R. W. Lawler
ARI 6 S 10
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

Dr. Alan M. Lesgold
Learning Research and
Development Center
University of Pittsburgh
Pittsburgh, PA 15260

Dr. Alan Leshner
Deputy Division Director
Behavioral and Neural Sciences
National Science Foundation
1800 G Street
Washington, DC 20550

Dr. Jim Levin
Department of
Educational Psychology
210 Education Building
1310 South Sixth Street
Champaign, IL 61820-6990

Dr. John Levine
Learning R&D Center
University of Pittsburgh
Pittsburgh, PA 15260

Dr. Clayton Lewis
University of Colorado
Department of Computer Science
Campus Box 430
Boulder, CO 80309

Matt Lewis
Department of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213

Library,
Naval War College
Newport, RI 02940

Library,
Naval Training Systems
Center
Orlando, FL 32813

Science and Technology Division,
Library of Congress
Washington, DC 20540

Dr. Jane Malin
Mail Code SR 111
NASA Johnson Space Center
Houston, TX 77058

Dr. Sandra P. Marshall
Dept. of Psychology
San Diego State University
San Diego, CA 92182

Dr. Humberto Maturana
University of Chile
Santiago
CHILE

Dr. Richard E. Mayer
Department of Psychology
University of California
Santa Barbara, CA 93106

Dr. James McBride
Psychological Corporation
c/o Harcourt, Brace,
Javanovich Inc.
1250 West 6th Street
San Diego, CA 92101

Dr. James L. McGaugh
Center for the Neurobiology
of Learning and Memory
University of California, Irvine
Irvine, CA 92717

Dr. Gail McKoon
CAS/Psychology
Northwestern University
1859 Sheridan Road
Kresge #230
Evanston, IL 60201

Dr. Joe McLachlan
Navy Personnel R&D Center
San Diego, CA 92152-6800

1987/03/09

Distribution List [Pittsburgh/Lesgold] NR 4422539

Dr. James S. McMichael
Navy Personnel Research
and Development Center
Code 05
San Diego, CA 92152

Dr. Barbara Means
Human Resources
Research Organization
1100 South Washington
Alexandria, VA 22314

Dr. Douglas L. Medin
Department of Psychology
University of Illinois
603 E. Daniel Street
Champaign, IL 61820

Dr. George A. Miller
Department of Psychology
Green Hall
Princeton University
Princeton, NJ 08540

Dr. Andrew R. Molnar
Scientific and Engineering
Personnel and Education
National Science Foundation
Washington, DC 20550

Dr. William Montague
NPRDC Code 13
San Diego, CA 92152-6800

Dr. Nancy Morris
Search Technology, Inc.
5550-A Peachtree Parkway
Technology Park/Summit
Norcross, GA 30092

Dr. Randy Mumaw
Program Manager
Training Research Division
HumRRO
1100 S. Washington
Alexandria, VA 22314

Dr. Allen Munro
Behavioral Technology
Laboratories - USC
1845 S. Elena Ave., 4th Floor
Redondo Beach, CA 90277

Chair, Department of
Computer Science
U.S. Naval Academy
Annapolis, MD 21402

Chair, Department of
Systems Engineering
U.S. Naval Academy
Annapolis, MD 21402

Technical Director,
Navy Health Research Center
P.O. Box 85122
San Diego, CA 92138

Dr. Allen Newell
Department of Psychology
Carnegie-Mellon University
Schenley Park
Pittsburgh, PA 15213

Dr. Mary Jo Nissen
University of Minnesota
N218 Elliott Hall
Minneapolis, MN 55455

Dr. A. F. Norcio
Computer Science and Systems
Code: 7590
Information Technology Division
Naval Research Laboratory
Washington, DC 20375

Dr. Donald A. Norman
Institute for Cognitive
Science C-015
University of California, San Diego
La Jolla, California 92093

Deputy Technical Director,
NPRDC Code 01A
San Diego, CA 92152-6800

Director, Training Laboratory,
NPRDC (Code 05)
San Diego, CA 92152-6800

Director, Manpower and Personnel
Laboratory,
NPRDC (Code 06)
San Diego, CA 92152-6800

1987/03/09

Distribution List [Pittsburgh/Lesgold] NR 4422539

Director, Human Factors
& Organizational Systems Lab,
NPRDC (Code 07)
San Diego, CA 92152-6800

Fleet Support Office,
NPRDC (Code 301)
San Diego, CA 92152-6800

Library, NPRDC
Code P201L
San Diego, CA 92152-6800

Technical Director,
Navy Personnel R&D Center
San Diego, CA 92152-6800

Commanding Officer,
Naval Research Laboratory
Code 2627
Washington, DC 20390

Dr. Harold F. O'Neil, Jr.
School of Education - WPH 801
Department of Educational
Psychology & Technology
University of Southern California
Los Angeles, CA 90089-0031

Dr. Michael Oberlin
Naval Training Systems Center
Code 711
Orlando, FL 32813-7100

Dr. Stellan Ohlsson
Learning R & D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

Office of Naval Research,
Code 1142BI
800 N. Quincy Street
Arlington, VA 22217-5000

Office of Naval Research,
Code 1142
800 N. Quincy St.
Arlington, VA 22217-5000

Office of Naval Research,
Code 1142PS
800 N. Quincy Street
Arlington, VA 22217-5000

Office of Naval Research,
Code 1142CS
800 N. Quincy Street
Arlington, VA 22217-5000
(6 Copies)

Psychologist,
Office of Naval Research
Branch Office, London
Box 39
FPO New York, NY 09510

Special Assistant for Marine
Corps Matters,
ONR Code 00MC
800 N. Quincy St.
Arlington, VA 22217-5000

Psychologist,
Office of Naval Research
Liaison Office, Far East
APO San Francisco, CA 96503

Dr. Judith Orasanu
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Douglas Pearse
DCIEM
Box 2000
Downsview, Ontario
CANADA

Dr. James W. Pellegrino
University of California,
Santa Barbara
Department of Psychology
Santa Barbara, CA 93106

Dr. Virginia E. Pendergrass
Code 711
Naval Training Systems Center
Orlando, FL 32813-7100

1987/03/09

Distribution List [Pittsburgh/Lesgold] NR 4422539

Dr. Nancy Pennington
University of Chicago
Graduate School of Business
1101 E. 58th St.
Chicago, IL 60637

Military Assistant for Training and
Personnel Technology,
OUSD (R & E)
Room 3D129, The Pentagon
Washington, DC 20301-3080

Dr. Steven Pinker
Department of Psychology
E10-018
M.I.T.
Cambridge, MA 02139

Dr. Martha Polson
Department of Psychology
Campus Box 346
University of Colorado
Boulder, CO 80309

Dr. Peter Polson
University of Colorado
Department of Psychology
Boulder, CO 80309

Dr. Michael I. Posner
Department of Neurology
Washington University
Medical School
St. Louis, MO 63110

Dr. Mary C. Potter
Department of Psychology
MIT (E-10-032)
Cambridge, MA 02139

Dr. Paul S. Rau
Code U-32
Naval Surface Weapons Center
White Oak Laboratory
Silver Spring, MD 20903

Dr. Lynne Reder
Department of Psychology
Carnegie-Mellon University
Schenley Park
Pittsburgh, PA 15213

Dr. James A. Reggia
University of Maryland
School of Medicine
Department of Neurology
22 South Greene Street
Baltimore, MD 21201

Dr. Wesley Regian
AFHRL/MOD
Brooks AFB, TX 78235

Dr. Fred Reif
Physics Department
University of California
Berkeley, CA 94720

Dr. Gil Ricard
Mail Stop C04-14
Grumman Aerospace Corp.
Bethpage, NY 11714

Dr. Linda G. Roberts
Science, Education, and
Transportation Program
Office of Technology Assessment
Congress of the United States
Washington, DC 20510

Dr. Paul R. Rosenbaum
Educational Testing Service
Princeton, NJ 08541

Dr. William B. Rouse
Search Technology, Inc.
5550-A Peachtree Parkway
Technology Park/Summit
Norcross, GA 30092

Dr. David Rumelhart
Center for Human
Information Processing
Univ. of California
La Jolla, CA 92093

Dr. Walter Schneider
Learning R&D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

1987/03/09

Distribution List [Pittsburgh/Lesgold] NR 4422539

Dr. Miriam Schustack
Code 51
Navy Personnel R & D Center
San Diego, CA 92152-6800

Dr. Marc Sebrechts
Department of Psychology
Wesleyan University
Middletown, CT 06475

Dr. Colleen M. Seifert
Intelligent Systems Group
Institute for
Cognitive Science (C-015)
UCSD
La Jolla, CA 92093

Dr. Ben Shneiderman
Dept. of Computer Science
University of Maryland
College Park, MD 20742

Dr. Robert S. Siegler
Carnegie-Mellon University
Department of Psychology
Schenley Park
Pittsburgh, PA 15213

Dr. Herbert A. Simon
Department of Psychology
Carnegie-Mellon University
Schenley Park
Pittsburgh, PA 15213

LTCOL Robert Simpson
Defense Advanced Research
Projects Administration
1400 Wilson Blvd.
Arlington, VA 22209

Dr. H. Wallace Sinaiko
Manpower Research
and Advisory Services
Smithsonian Institution
801 North Pitt Street
Alexandria, VA 22314

Dr. Richard E. Snow
Department of Psychology
Stanford University
Stanford, CA 94306

Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Kathryn T. Spoehr
Brown University
Department of Psychology
Providence, RI 02912

Dr. James J. Staszewski
Research Associate
Carnegie-Mellon University
Department of Psychology
Schenley Park
Pittsburgh, PA 15213

Dr. Robert Sternberg
Department of Psychology
Yale University
Box 11A, Yale Station
New Haven, CT 06520

Dr. Kurt Steuck
AFHRL/MOD
Brooks AFB
San Antonio TX 78235

Dr. Paul J. Sticha
Senior Staff Scientist
Training Research Division
HumRRO
1100 S. Washington
Alexandria, VA 22314

Dr. John Tangney
AFOSR/NL
Bolling AFB, DC 20332

Dr. Kikumi Tatsuoka
CERL
252 Engineering Research
Laboratory
Urbana, IL 61801

Dr. Perry W. Thorndyke
FMC Corporation
Central Engineering Labs
1185 Coleman Avenue, Box 580
Santa Clara, CA 95052

1987/03/09

Distribution List [Pittsburgh/Lesgold] NR 4422539

Dr. Sharon Tkacz
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Douglas Towne
Behavioral Technology Labs
1845 S. Elena Ave.
Redondo Beach, CA 90277

Headquarters, U. S. Marine Corps
Code MPI-20
Washington, DC 20380

Dr. William Uttal
NOSC, Hawaii Lab
Box 997
Kailua, HI 96734

Dr. Kurt Van Lehn
Department of Psychology
Carnegie-Mellon University
Schenley Park
Pittsburgh, PA 15213

Dr. Beth Warren
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02138

Dr. Keith T. Wescourt
FMC Corporation
Central Engineering Labs
1185 Coleman Ave., Box 580
Santa Clara, CA 95052

Dr. Douglas Wetzel
Code 12
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Barbara White
Bolt Beranek & Newman, Inc.
10 Moulton Street
Cambridge, MA 02238

Dr. Christopher Wickens
Department of Psychology
University of Illinois
Champaign, IL 61820

Dr. Heather Wild
Naval Air Development
Center
Code 6021
Warminster, PA 18974-5000

Dr. Robert A. Wisher
U.S. Army Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Martin F. Wiskoff
Navy Personnel R & D Center
San Diego, CA 92152-6800

Dr. Dan Wolz
AFHRL/MOE
Brooks AFB, TX 78235

Dr. Wallace Wulfeck, III
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Joe Yasatuke
AFHRL/LRT
Lowry AFB, CO 80230

Dr. Joseph L. Young
Memory & Cognitive
Processes
National Science Foundation
Washington, DC 20550